Rational Drug Combinations Design against Intratumoral Heterogeneity and Clonal Evolution

By

Boyang Zhao

B.S.E. in Biomedical Engineering
University of Michigan, Ann Arbor, 2011

Submitted to the Computational and Systems Biology Program in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Computational and Systems Biology
at the
Massachusetts Institute of Technology

February 2016

© 2016 Massachusetts Institute of Technology. All Rights Reserved.

Signature of Author

... Computational and Systems Biology Graduate Program
December 30, 2015

Certified by

Michael T. Hemann
Thesis Supervisor
Associate Professor of Biology

Certified by

Douglas A. Lauffenburger
Thesis Supervisor
Ford Professor of Biological Engineering, Chemical Engineering, and Biology

Approved by

Christopher B. Burge
Professor of Biology and Biological Engineering
Director, Computational and Systems Biology Graduate Program
Rational Drug Combinations Design against Intratumoral Heterogeneity and Clonal Evolution

By

Boyang Zhao

Submitted to the Computational and Systems Biology Program on December 30, 2015
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computational and Systems Biology

Abstract

Cancer is a clonal evolutionary process. This results in complex clonal architecture and intratumoral heterogeneity in each patient. This also presents challenges for effective therapeutic intervention—constant selective pressure to induce or select pre-existing resistant subclones toward drug resistance. Mathematical/computational modeling from population genetics, evolutionary dynamics, and engineering are being utilized to a greater extent in recent times to study tumor progression, intratumoral heterogeneity, drug resistance, and rational drug scheduling/combinations design. In this thesis we present several joint quantitative and experimental approaches for the rational design of drug combinations to tackle the issue of intratumoral heterogeneity and clonal evolution.

Using a tractable experimental system with pre-defined tumor compositions, we derived computational approaches to rationally design drug combinations with the goal of minimizing a given heterogeneous tumor. We found that the best drug combinations can oftentimes be non-intuitive as they do not contain component drugs most effective for the individual subpopulations. This was the result of a need for combinatorial considerations on the effects of each drug on all subpopulations, hence at times leading to non-intuitive drug regimens. We validated our computational model predictions in vitro and in vivo in a preclinical model of Burkitt’s lymphoma, with predictable evolutionary trajectories upon treatment. Next, we extended this methodology to study the effects of more complex tumor heterogeneity on combinatorial drug design, with similar conclusions. Sampling and statistical analyses over a range of tumor compositions can further inform effective drug combinations under some uncertainty in initial tumor heterogeneity. Moving beyond a model where we have control of initial tumor composition, we sought to examine collateral resistance and sensitivity during clonal evolution. Using a murine model of Ph+ acute lymphoblastic leukemia, we performed drug selection and pharmacological screen experiments. We observed important evolutionary processes of selection and drift in giving rise to resistance to clinically used BCR-ABL1 inhibitors. Remarkably, the resistant population also became hyper-sensitized to non-classical BCR-ABL1 inhibitors at intermediate stages of the clonal evolution, in this so-called ‘temporally collateral sensitivity’. Mathematical modeling and experimentation brought additional insight into the evolutionary dynamics and mechanism of action, with demonstrated in vivo efficacy.

These quantitative approaches, complemented with extensive experimentation facilitate a principled approach to our understanding of making forward predictions that directly inform therapeutic drug regimen designs.

Supervisor: Michael T. Hemann, PhD
Title: Associate Professor of Biology

Supervisor: Douglas A. Lauffenburger, PhD
Title: Ford Professor of Biological Engineering, Chemical Engineering, and Biology
Acknowledgements

There are many people I have to thank throughout my years at MIT. Especially I thank my advisors Dr. Michael Hemann and Dr. Douglas Lauffenburger for their guidance, advice, support, and patience. It has also been a privilege to have Dr. Piyush Gupta and Dr. Dane Wittrup on my thesis committee and I am grateful for their time and feedback throughout this process. I also thank Dr. Franziska Michor for graciously agreeing to be on the thesis committee for my defense.

I thank all members of the Hemann lab, past and present, including Justin Pritchard, Luke Gilbert, Peter Bruno, Christian Braun, Eric Bent, Silvia Fenoglio, Nina Fenouille, Alex Puissant, Simona Dalin, and many others who made the lab environment fun and intellectually challenging. I am also very grateful for all members of the Lauffenburger lab for the discussions, advices, and support throughout these years, and my officemates, past and present, including Aaron Meyer, Jackie Shepard, Elizabeth Proctor, Marianna Sofman, and Kasper Renggli. I also thank our lab managers/administrators, Hsinhwa Lee, Ryan Hayman, and Holly Criscione, for keeping the labs running smoothly. I was also fortunate to have mentored my rotation student Joe Sedlak and undergraduate student Daphne Sun in the last few years.

I would also like to thank Dr. Chris Burge and Jacquie Carota for overseeing and maintaining the CSB program and in organizing all the events for an enriched learning/research community. My CSB classmates Max Wolf, Kelly Brock, Sarah Spencer, Thomas Willems, Asa Adadey and many others who have made this program special – and it has also been a delight to catch up during the CSB retreats or at random times on-campus. Lastly, I thank my family, friends, and past mentors who supported me and guided me in one way or another onto this path.
# Table of Contents

**Chapter I** Introduction................................................................................................................... 5  
Tumor Clonal Evolution and Intratumoral Heterogeneity ................................................................. 6  
Quantitative Approaches to Modeling Clonal Evolution ................................................................. 7  
Fitness Landscapes.......................................................................................................................... 12  
Traversing on the Fitness Landscape.............................................................................................. 14  
Effects of Drug Treatment on Clonal Evolution ............................................................................ 16

**Chapter II** Addressing genetic tumor heterogeneity through computationally predictive combination therapy .................................................................................................................. 25

**Chapter III** Intratumor heterogeneity alters most effective drugs in designed combinations..... 57

**Chapter IV** Exploiting temporal collateral sensitivity in tumor clonal evolution ...................... 103

**Chapter V** Concluding Remarks .................................................................................................. 155
Chapter I

Introduction

Portions submitted (with modifications) as: Zhao, B., Hemann, M. T., & Lauffenburger, D. A. Modeling tumor clonal evolution for drug combinations design. Under review.
Recent data from tumor sequencing has increased attention on the broad relevance of intratumoral heterogeneity in cancer patients and their treatment. In light of these studies, the tumor biology field now more than ever regards cancer as an ongoing evolutionary process. As such, an important opportunity in understanding this malignancy and its dynamics, as influenced by selective pressures from therapeutic intervention, will require capitalizing on quantitative methods from population genetics, evolution, and engineering. A number of excellent reviews on tumor heterogeneity (1-7), drug resistance (8-11), drug combinations design (12-15) are available, so we aim here to integrate germane quantitative approaches, with a focus on recent developments and applications to understanding and predicting therapeutic effects. Whenever possible, we will refer readers to relevant existing reviews, but will also provide brief historical context, and most importantly will offer introduction to foundational mathematical frameworks.

**Tumor Clonal Evolution and Intratumoral Heterogeneity**
The notion of cancer as a clonal evolutionary process dates to seminal work in 1976 by Nowell (16). A key consequence of tumor clonal evolution is intratumoral heterogeneity – as the founder clone develops successive alterations with fitness advantages subject to selection forces (e.g., tumor progression, metastasis, and drug resistance). Heterogeneity in tumor cells across different regions was indeed observed by pathologists as early as the 1800s, based on cell morphology and other cytological characteristics (7). Cytogeneticists have since used chromosome G-banding, spectral karyotyping, and fluorescence in situ hybridization to further show distinct subpopulations with copy number variations and chromosomal rearrangements (7). More recently, with the advent of micro-technologies (e.g., SNP arrays, comparative genomic hybridization microarrays) and next-generation sequencing (NGS), the field has begun to realize at much greater resolution the complexity of tumor heterogeneity and clonal dynamics.

Clonal evolution may be conceptualized as progressing via accumulation of what many term ‘driver’ and ‘passenger’ mutations. From an evolutionary perspective, drivers can be considered as genetic (or epigenetic) alterations that offer significant relative fitness advantage, while passengers are mutations appearing to present only weak or neutral influences. Numerous computational techniques have been developed to test for positive selection and identification of such ‘drivers’ (17). Over time, evolutionary processes – genetic drift and Darwinian selection – play a critical role in the progression and fixation (i.e., establishment) of subclones with these various alterations. Specifically, if the supply rate of driver mutations is slower than the time it takes for the established subclone to selectively sweep through the population, a classical linear model of clonal evolution (6, 18) is followed. In contrast, however, experimentally observed tumor clonal architectures frequently are found instead to represent a complex branched clonal process, as originally described by Nowell: multiple subclones may present and compete
via clonal interference prior to fixation (6, 18, 19). Such branched clonal evolutionary processes have been observed in many tumor types, including leukemias (20–22), renal-cell carcinoma (23), pancreatic cancer (24, 25), amongst others. Recently a ‘big bang’ model of tumor evolution was observed in an analysis of 349 glands from 15 colorectal tumors – where a majority of alterations transpire early during transformation, resulting in tumor growth via concomitant single expansion of multiple mixed subclones without selective sweeps and selection pressures (26). It should be emphasized, moreover, that heterogeneity is not restricted to genetic alterations, as epigenetic heterogeneities have also been reported (27).

Intratumoral heterogeneity may have functional roles beyond mere indication of tumor history. Studies in breast cancer xenograft models with component or combined mixed heterogeneous tumor populations suggest the importance of heterogeneity in tumor maintenance (28). Drug treatment also can impose tremendous selective pressure on reshaping the clonal architecture, as will be discussed later in this chapter.

**Quantitative Approaches to Modeling Clonal Evolution**

Mathematical modeling of tumor development and metastasis has been the subject of comprehensive reviews periodically over the past decade (9, 29, 30). Building on these, here we will introduce vital fundamental tools from population genetics (31, 32) and evolutionary dynamics (33) as applied to cancer and then move to emphasize a view of clonal evolution based on fitness landscapes and focus on relation to therapy. Mathematical studies of cancer began as early as the 1950s, from works by Nordling (34), Armitage and Doll (35, 36), Fisher (37), and Knudson (38) on the age-incidence of cancer and the multistage theory of carcinogenesis. Since then, a large number of models have been based on well-mixed populations with the assumption of no cell-cell interaction. These models provide a tractable system to study cancer dynamics and serve as the foundation for additional model complexity. Several important evolutionary forces (e.g., genetic drift, selection, and mutation) underlie much of our observed tumor dynamics and are of differential importance under specific parameter regions (e.g., population size). In particular, in large population size, stochastic fluctuations can be negligible, and as such population dynamics can be studied using deterministic models. Here, the same outcome will generally be achieved given the same initial conditions. On the other hand, in a stochastic process, the same initial condition can yield different outputs (as described by probability distributions) – a feature that is important in modeling processes such as genetic drift.
Stochastic models

Several tools from stochastic modeling have been utilized to analyze tumor evolution. These include Markov models, where a system’s future state depends only on the current state of the system (i.e., the Markov property) (39). Specifically, a Markov chain consists of a series of probabilistic variables, satisfying the Markov property, with the transition between states described by the transition matrix of rate constants for movement of the variables between different values. A Markov chain is considered as a time-homogeneous process if the transition matrix stays the same after each step; otherwise the process is time-nonhomogeneous. A Markov process can be further classified based on it being cast as discrete or continuous with respect to time, and on the ‘space’ of the system’s state being characterized by variables either discrete or continuous in nature. We provide a brief overview of the Markov models that have been used, which can be broadly considered within two categories: finite constant and finite fluctuating population size models.

Assuming, on one hand, a constant population size, the Moran process and Fisher-Wright process (Fig. 1) have been classical and standard models in population genetics to study genetic drift (and hence changes in allele frequency over time in a population, with possible incorporation of selection and/or mutation).

A Moran process is a discrete-time Markov chain (specifically a birth/death process) (33) with overlapping generations. We suppose a total fixed population size $N$ and two cell types (A and B) and corresponding fitness $f_A$ and $f_B$. At each generation (time step), a random cell is selected to replicate with probability proportional to its fitness, and another cell uniformly selected to die. As such, this constitutes a birth/death pair event with constant overall population size and two absorbing states ($i = 0$ or $i = N$). Under neutral drift (where $f_A = f_B$), the fixation probability of cell type A with $i$ cells is given by,

$$x_i = \frac{i}{N}$$

Under selection with relative fitness $r$ (i.e. suppose cell type B has fitness of 1 and cell type A has fitness of $r$), the fixation probability of cell type A with $i$ cells is given by,

$$x_i = \frac{1 - 1/r^i}{1 - 1/r^N}$$

Mutation can also be incorporated, such that with a mutation rate $u$, the rate of evolution from a shift of the entire population being A to B is given by

$$R = Nux_1$$
Moran process models have been used in a number of studies on tumor initiation and progression (40–42), in understanding the incidence rates (and their dependence on fitness distributions), rates of allele inactivation/activation, and stochastic tunneling.

In a Wright-Fisher process, the entire population is sampled at each generation (instead of just a single birth/death pair event), and as such assumes non-overlapping generations. To stay with the two-cell type example above, a Wright-Fisher process here involves given a total fixed population size $N$, we sample with replacement from the current population pool to derive the population composition at the next generation. Thus, the probability of cell type $A$ with $j$ cells in the current generation to contain $i$ cells at the next generation simply follows the binomial distribution,

$$
P(N_j \rightarrow N_i) \sim \text{Bin}\left(i; N, \frac{j}{N}\right) = \binom{N}{i} \left(\frac{j}{N}\right)^i \left(1 - \frac{j}{N}\right)^{N-i}
$$

The association between the two processes is apparent, as one generation in the Wright-Fisher process corresponds to $N$ generations in the Moran process. In the limit of large population size and with rescaling, both models lead to similar results (e.g. fixation probabilities) (31, 43). Other evolutionary forces such as selection and mutation can also be incorporated, and with multi-type subpopulations, in which case a multinomial distribution would be used. Although Moran model has more benefits with regards to exact analyses, Wright-Fisher process is much more computationally efficient, particularly for large population sizes. It is also of note that Wright-Fisher process and Moran process have been used with modifications for an exponential growing tumor population size (44, 45). The Wright-Fisher process has also been used in the study of tumor progression (44, 46), clonal interference (47), and intratumoral heterogeneity (48).

With a finite fluctuating population size, branching process (49) has been widely used for the study of tumor evolution and dynamics with exponential growth (Fig. 1). The basic premise is that each cell in a population independently gives rise to a random number of offspring accordingly to a probability distribution at some later time. The simplest branching process is that of a single-type discrete-time Bienaymé-Galton-Watson process, which can be defined as,

$$
X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)}
$$
where \( X_n \) refers to the population size at the \( n \)th generation and each \( \zeta_j^{(n)} \) is an independently and identically distributed random variable describing the offspring distribution for the \( j \)th cell at the \( n \)th generation. Probability generating functions are the primary tool used in the analysis of these processes.

Several special cases exist. If Galton-Watson process follows a Poisson offspring distribution conditioned at a constant population size, we can obtain the Wright-Fisher process as described above (49). If each cell divides (into two offspring) or dies with a specific birth or death rate, then this branching process is a nonhomogeneous continuous-time birth/death process. Note that birth/death processes are a special case of continuous-time Markov chains. In general, continuous-time Markov chains can be simulated with the Gillespie algorithm (or with modified methods such as tau-leaping for computational efficiency).

**Figure 1. Illustrations of selected stochastic processes.** A Moran process is a birth/death process for which at each generation, a cell is chosen to divide and a cell is chosen to die. In a Wright-Fisher process, the entire population is replaced at the next generation \( t + 1 \), via sampling with replacement from the population at generation \( t \). While both Moran and Wright-Fisher maintains a finite population size, a branching process models fluctuating population sizes. Shown in figure is a discrete time branching process, where at each generation, each cell either divides or dies.
Branching process has been used in the study of tumor heterogeneity and initiation/progression (9, 50–56). For example, Bozic et al. (50) studied the accumulation of driver and passenger mutations and the timing of these events in relation to epidemiologic and clinical observations. Parameterization to experimental data suggests that the selective advantage conferred by somatic mutations is very small. This model was later extended with consideration to epistasis (55). Studies by birth/death models were also applied to studying the role of moderate deleterious passenger mutations on tumors and their effects in potentially alternating cancer progression and therapeutic response (57, 58). These mathematical models have also becoming increasingly used in studying effects of drug treatment (as further discussed in a later section).

An alternative to analyze stochastic processes that can be more tractable is with approximation using diffusion models. **Fokker-Planck** (Kolmogorov forward) equation models the time evolution of a probability distribution density \( \varphi(p, t) \) for allele frequency \( p \) and time \( t \) of a Markov process as,

\[
\frac{\partial}{\partial t} \varphi(p, t) = -\frac{\partial}{\partial p} M(p) \varphi(p, t) + \frac{1}{2} \frac{\partial^2}{\partial p^2} V(p) \varphi(p, t)
\]

where \( M(p) \) and \( V(p) \) represents the mean and variance of the Markov process, and models directional (e.g. selection) and non-directional (e.g. genetic drift) processes, respectively. The diffusion approximations are helpful in determining fixation probabilities and mean fixation times (43). Recently, Tomasetti et al. used the diffusion approximation to the Moran process to estimate the expected number of passenger mutations, and their studies suggest that a substantial number of somatic mutations in tumor may have occurred prior to tumor initiation (59).

**Deterministic models**

At large population size, the population can be usefully modeled using deterministic models with difference or differential equations. In evolutionary game dynamics, the quasispecies equation (formulated by Eigen and Schuster for the study of molecular evolution) have been used to describe selection and mutation of an infinitely large population (33). Given \( x_i \) as the frequency of genotype \( i \) with fitness \( f_i \) and a mutation matrix \( Q \), we can describe the derivative of \( x_i \) as,

\[
\dot{x}_i = \sum_j x_j f_j q_{ji} - \phi(x)x_i
\]

where \( \phi(x) \) is average population fitness and defined as,

\[
\phi(x) = \sum_j x_j f_j
\]
The quasispecies equation is a special case of the replicator-mutator equation, which also accounts for frequency-dependent selection (where each fitness value \( f_i \) depends on the frequency of other cells in the population). Page and Nowak have provided a review on the relation among these different deterministic evolutionary dynamics models (e.g., quasi-species equation, replicator-mutator equation, replicator equation, Lotka-Volterra equation, Price equation, etc.) (60).

Deterministic models (61) and hybrid of deterministic/stochastic models (62) have been used in the study of CML dynamics under treatment. A wide range of differential equation based models has also been used for the study of drug scheduling (discussed later). Lastly, the quasi-species equation was used in the study of genetic instability during tumor progression (63).

**Fitness Landscapes**

The mathematical models discussed thus far lack connection of genotype to phenotype, and of either or both to parameters characterizing ‘fitness’ of the population under a given environment (e.g., drug treatment condition). This can be accomplished via description of a fitness landscape (also known as adaptive landscape) (64, 65) – a mapping of multidimensional genotype (and/or phenotype) space to its corresponding fitness. An idealized realization of this space may be seen in three dimensions (Fig. 2A), with the xy plane representing the genotype space and the z-axis representing fitness. Fitness landscape was originally proposed by Wright in a seminal 1932 paper (66) on ‘shifting balance’ theory, arguing that population evolution toward higher fitness (a high adaptive peak) occurs in three phases: genetic drift ‘shifts’ a subpopulation from one adaptive valley to another valley; subsequent selection drives the subpopulation to follow the fitness gradient from the valley base toward the specific adaptive peak; and different adaptive peaks compete, with the fitter peak eventually become populated by the entire population. Later advances on understanding evolutionary adaptation followed (reviewed elsewhere (63)) emphasized a molecular basis of inheritance, first by Smith’s analysis of protein space (67).

Subsequently, Gillespie tried to investigate specifically the distribution of fitness values by proposing that the fitness values are likely to be drawn from the extreme tail of the distribution, such that extreme value theory (EVT) can be applied to study the tails of distribution independent of knowing the distribution itself (68).
Several approaches exist for modeling fitness landscapes (64), and broadly fall under the following categories: random field, sequence-structure, and phenotype-fitness models. **Random field models** (64, 71) includes the Kauffman's NK model (72), which for a system with $N$ total number of loci and $K$ number of interactions between loci, the fitness for particular genotype $x$ is defined as,

$$f(x) = \sum_{i=1}^{N} f_i(x_i, x_1, ..., x_K)$$

where $f_i$ is a fitness function (also known as fitness contribution) and defined as an independent and identically distributed random variable. The choice of the $K$ interaction partners can be chosen at random uniformly. By tuning the $N$ and $K$ parameter values, one can achieve fitness landscapes with different ruggedness. When $K = N - 1$, the genotypes become uncorrelated and is known as the House of Cards (HoC) model (73). When $K = 0$, the landscape is additive and non-epistatic and is known as Mount Fuji model. The Rough Mount Fuji (RMF) model (74) is an additive landscape with added noise. **Sequence-structure models** (75) describe explicit mapping of sequence or structure (e.g. RNA, DNA, protein) to
predicted fitness metrics (e.g. stability, affinity, etc). Lastly, **phenotype-fitness models** include heuristic phenotype-based landscapes loosely based on Fisher's geometric model (65).

Aside from these mathematical models of fitness landscapes, there has been an increasing number of empirical landscapes derived based on evolutionary or systematic combinatorial experiments (reviewed elsewhere (64, 71)). Recent developments using CRISPR/Cas9 have enabled saturation editing and analyses at the endogenous locus (76), specifically applying this to genomic regions of BRCA1 and DBR1 and measured the functional impact/fitness of different alterations. It is conceivable that this technology will enable the detailed investigation in a native context the mapping between genotype and fitness.

**Traversing on the Fitness Landscape**

Topology of landscapes is of particular importance as it provides information regarding evolutionary trajectories, predictability, rate of adaptation, etc. In particular, rugged landscapes (bearing multiple peaks) can occur as a result of sign epistasis, whereby the effects of a specific allele depend contextually on the genetic background at other loci. This would cause certain paths along the fitness landscape to become inaccessible. Pathway inaccessibility in a rugged landscape can constrain evolutionary trajectories and consequently increases repeatability and predictability. Experimental evidence based on mutagenesis and fitness measurements (viz. drug resistance) suggest evolution may be constrained with a large number of inaccessible mutational trajectories (77, 78). Additionally, several experimental studies also showed convergent and parallel evolution when the populations (E. coli and S. cerevisiae) were independently evolved (79–81). Evolutionary predictability and convergent evolution is of particular relevance as to our understanding of whether tumor clonal evolution has defined trajectories and potential forward predictions on how tumor responds to treatment (as further discussed in next section).

For analysis of evolutionary predictability, several parameters must be considered (64, 82): mutation rate $\mu$, selection coefficient $s$, and population size, $N$. With a small population size, this is under the strong-selection-weak-mutation (SSWM) regime, a condition considered and studied by Gillespie in his work on EVT and mutational landscapes. SSWM assumes the mutation supply rate is sufficiently low ($N\mu \ll 1$; rendering negligible more than one mutation event) and selection is high ($Ns \gg 1$; yielding mutations with strongly beneficial or deleterious effects). This suggests that predictability is low as stochasticity and subsequent fixation dominates. With increasing population size, the population is less unlikely to be monomorphic, and SSWM dynamics transitions to a ‘greedy’ adaptation dynamics, where clonal inference occurs and bias the selection of the fitter clones; correspondingly, predictability of system
outcome is enhanced. At sufficiently large population sizes clones possessing multiple mutations may be present, allowing passage along trajectories not previously accessible (i.e., enables valley crossing via ‘tunnels’ through intermediate states). This then generates greater stochasticity and accordingly decreased predictability (82). Furthermore, a trade-off exists between the rate of adaptation and eventual fitness level achieved, whereby small population sizes can reach higher fitness (as stochasticity allows exploration of more diverse areas of the landscape), whereas larger population sizes can remain trapped at the local fitness peak (83–85).

The concept of ‘evolutionary trap’ (Fig. 2B) has been studied in microbial systems (e.g., S. cerevisiae, C. albicans) whereby karyotypically heterogeneous populations were driven by a first drug to a region of the fitness landscape with reduced diversity, followed by maximized kill using a second drug (69). Specific regions of the fitness landscape with distinct vulnerability under different conditions have also recently exploited using a concept of ‘temporal collateral sensitivity’ (Fig. 2C). Collateral sensitivity is a form of synthetic lethality, again studied frequently in bacterial evolution, for which due to evolutionary constraints and trade-offs, populations becoming resistant to one drug can be sensitized to other classes of drugs (86–89). The difference between Panel C and Panel B in Figure 2 is with respect to chronology; Panel B represents fixation of a subpopulation to a final state (that also happens to be resistant to drug A in this case) that is sensitive to drug C, whereas Panel C represents treatment with drug D during a transient period as a subpopulation is starting to develop resistance to drug A but before establishment of the final emergent clone.

New work has begun to introduce this notion of temporal collateral sensitivity (Fig. 2C) (as described in Chapter IV of this thesis) to understanding tumor evolution and drug responses, and especially to elucidating more effective sequential combination treatments. In a series of drug selection and small molecule screening experiments using murine Ph+ acute lymphoblastic leukemia cells, we have shown that an intermediate stage during clonal evolution toward resistance to classical BCR-ABL1 inhibitors can be hyper-sensitized to non-classical BCR-ABL1 inhibitors (70). This finding demonstrates potential opportunities on the utility of diverse small molecules for targeting distinct stages of clonal evolution as tumor transverse along a fitness landscape toward resistance.

It should be noted that although the general assumption is that of a static landscape, actual evolution may also involve an ever-changing fitness landscape that varies with clonal evolution (90, 91). The advent of next-generation sequencing has increased capabilities for high-resolution tracking of clonal evolution. In one study, high-resolution barcoding libraries (~500,000 barcodes) and sequencing enabled lineage
tracing of *S. cerevisiae* evolution and evaluation of the dynamics of mutations and corresponding fitness (92). In another evolution study using yeast, fluorophore-labeled competition assays combined with sequencing and mathematical modeling based on branching process models also explored the role of polyploidy on evolutionary adaptation (93).

**Effects of Drug Treatment on Clonal Evolution**

In parallel with the rising utility of NGS in studying intratumoral heterogeneity and tumor progression, analyses of matched biopsies of patients prior and post drug treatment has also revealed extensive clonal dynamics. The mechanisms by which resistance/relapse occurs can be via (i) *de novo* mutations (e.g. genotoxic chemotherapy that induces mutagenesis), (ii) selection of pre-existing resistant subclone with higher fitness, or (iii) tumor reduction and competitive release (whereby significant tumor reduction leads to a regime of small population size, followed by the outgrowth of a different subclone) (2). Experimental and clinical evidence has been extensive and reviewed elsewhere (2). Of central importance are determination in many cases of pre-existing subpopulations that ultimately expand to dominance at relapse (94, 95). Here, we will elaborate on the quantitative approaches taken (using many tools we have discussed in earlier sections) to interrogate the effects of therapeutic intervention through an evolutionary lens.

**Empirical drug combinations design**

Much of the early clinical work on empirical combination chemotherapy design was initiated by Frei and Freireich in the early 1950s. These studies have led to some of the common drug regimens in existence even to this day. With the advent of targeted therapy (and now advances in immuno-oncology), we have a greater arsenal of drugs to attempt to rationally combine. Moreover, we can design not only based on non-overlapping toxicity, but also on mechanism of actions and resistance (12, 13). We will further emphasize this trend in light of other converging conceptual advances and methodologies discussed in our concluding remarks in Chapter V of this thesis.

**Stochastic and deterministic models**

Early quantitative work on resistance was initiated by Luria and Delbrück in their classic phage resistance studies (96) in 1943. The work demonstrated the use of fluctuation analysis to confirm the emergence of resistance via Darwinian selection and expansion of a pre-existing bacterial subpopulation, rather than via induction. Assuming deterministic exponential growth for both wild-type and mutant populations and probabilistic mutation, they derived a distribution (now known as the Luria-Delbrück distribution (97)) describing the number of resistant colonies. Tumor growth kinetics under therapy was studied soon after
using murine L1210 leukemia cells, by Law (for antimetabolite resistance via selection and also showed superiority of drug combinations over single drug treatment) (98, 99) and by Skipper (for the Skipper-Schabel-Wilcox model on exponential tumor growth and log-kill tumor regression) (100). Norton and Simon (101, 102) subsequently observed in the 1970s that some tumors follow a Gompertz growth model, facilitating straightforward prediction using differential equation models that sequential dose dense drug combination schedule can be superior over alternating treatment. This result arises from the former treatment typically having higher dose density along with integrated drug effects. accordingly, dose intensification would prevent rapid leukemia cell regrowth under Gompertz growth kinetics. In the 1980s, Goldie and Coldman (103–105) applied stochastic mathematical models (of growth dynamics and incorporated mutation) to predict that the most efficacious regimen is an alternating drug combination schedule that is given as early as possible to minimize chances of resistance. The concept of dose intensity has also been explored using an empirically-derived equation to calculate a summation dose-intensity (SDI) of drug combinations, proposed by Hryniuk (106, 107).

Many of the mathematical models have since been extended and the issue of drug scheduling has been studied extensively (reviewed elsewhere (8–10, 30)). In recent studies, a linear quadratic model was used to derive optimal radiotherapy scheduling for glioblastoma treatment, with demonstrated improved survival on mice (108). The concept of ‘adaptive therapy’ was also recently proposed based on mathematical modeling to show potential efficacy in a modulated drug scheduling that controls a stable tumor burden, allowing a sensitive tumor subpopulation to suppress the outgrowth of the resistant subpopulation (109). The differential fitness differences between resistant and sensitive subpopulations have been studied in other tumor models that provide rationales for drug holiday schedules. Here, if the resistant subpopulation has relative lower fitness in the absence of drug, this suggests that intermittent drug dosing will enable the alternating rise and fall of the wild-type and resistant subpopulations, thus delaying resistance onset. This idea has been analyzed using stochastic birth/death process modeling with experimental validation in a non-small cell lung cancer (NSCLC) model in vitro and in vivo (110). This was also corroborated in a recent experimental study in BRAF(V600E) melanoma models, where due to a fitness defect of the resistant subpopulation in the absence of drug, an intermittent dosing schedule was shown to be effective (111).

In addition to drug scheduling, mathematical models have also been used in determining (and quantifying) the existence of resistant subpopulations prior to treatment. Due to the limit of detection of NGS technologies (at 0.1-1% due to its error rate) (112), detection of rare minor subpopulation is oftentimes a challenge. Although sophisticated barcoding methods have been developed with significant
lower error rates (113), these have yet to be widely adopted and implemented. However, several studies utilized mathematical modeling parameterized based on experimentally derived tumor dynamics data to indicate and estimate the minor subpopulation. Diaz et al. employed a branching process model to conclude the existence of rare KRAS-mutant cells prior to initiation of monotherapy with panitumumab in patients with colorectal cancer, based on observed clinical kinetics of resistance from circulating tumor DNA (ctDNA). This was similarly used in estimating the number of pre-existing BCR-ABL1 resistant mutants in vitro (70, 114). Other studies have similarly begun to address these questions in regards to pre-existence and risk of relapse (115–119).

**Optimal control theory and combinatorial optimization**

Besides the stochastic and deterministic models described earlier, complementary quantitative approaches from engineering disciplines have been applied to drug combinations and drug scheduling since the 1970s. Specifically, with dynamical systems modeled using deterministic differential equations, we can apply many concepts from optimal control theory in engineering to study the stability and control of cancer – with the goal of maximizing tumor kill and minimizing drug toxicity. Bahrami and Kim (120) first reported applications of optimal control theory to minimize tumor growth (and used an example with vincristine against murine AKR leukemia). Swan and Vincent (121) subsequently studied optimal control of chemotherapy in human IgG multiple myeloma. Since then, numerous studies have appeared (122, 123) (with many based on the control model formulated by Martin (124)) and have been solved analytically, using numerical methods (e.g. via formulation as a linear or nonlinear programming problem) or other heuristics methods (e.g. genetic algorithms, particle swarm optimization, simulated annealing, etc).

Optimization is pervasive in operations research, engineering design, and finance (e.g., as used in modern portfolio theory). Some of the aforementioned studies have also formulated the model as multi-objective optimization problems. Multi-objective optimization (a multi-objective decision making (MODM) problem) is one of two classes of multi-criteria decision making (MCDM) problems – the other being multi-attribute decision making (MADM). In multi-objective optimization, because of multiple competing objectives, the solution is a solution set (or Pareto optimal set) instead of a single optimal solution. An example is the Pareto frontier as used in production possibility curves in economics. Approaches for solving multi-objective optimization can be broadly categorized as no articulation, a priori articulation, a posteriori articulation, and progressive articulation (125).
While the majority of the mathematical optimization contributions have centered on drug scheduling, analogous efforts have also been made to study the choice of drug combinations. These include theoretical analysis using integer programming to determine the minimal set of drugs that maximizes the coverage of molecular target variants based on drug-target binding affinity properties (126). Recently, combinatorial (Chapter II of this thesis) and multi-objective optimization (Chapter III of this thesis) has also been applied to study how to design drug combinations in the context of intratumoral heterogeneity (127, 128). The results (with in vitro and in vivo validation in a Burkitt’s lymphoma model) show that consideration of heterogeneity can at times lead to non-intuitive optimal drug combinations.
REFERENCES

progression, 08018, 11 (2011).
74. T. Aita et al., Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: application to prolyl endopeptidase and thermolysin., Biopolymers 54, 64–79 (2000).
82. I. G. Szendro, J. Franke, J. A. G. M. de Visser, J. Krug, Predictability of evolution depends


107. W. Hryniuk, E. Frei, F. A. Wright, A single scale for comparing dose-intensity of all


Chapter II

Addressing genetic tumor heterogeneity through computationally predictive combination therapy


Supplemental Text and Tables are available online.
ABSTRACT
Recent tumor sequencing data suggests an urgent need to develop a methodology to directly address intra-tumor heterogeneity in the design of anti-cancer treatment regimens. We use RNA interference to model heterogeneous tumors, and demonstrate successful validation of computational predictions for how optimized drug combinations can yield superior effects on these tumors both in vitro and in vivo. Importantly, we discover here that for many such tumors knowledge of the predominant subpopulation is insufficient for determining the best drug combination. Surprisingly, in some cases the optimal drug combination does not include drugs that would treat any particular subpopulation most effectively, challenging straightforward intuition. We confirm examples of such a case with survival studies in a murine pre-clinical lymphoma model. Altogether, our approach provides new insights concerning design principles for combination therapy in the context of intratumoral diversity, data that should inform the development of drug regimens superior for complex tumors.

STATEMENT OF SIGNIFICANCE
This study provides the first example of how combination drug regimens, using existing chemotherapies, can be rationally designed to maximize tumor cell death, while minimizing the outgrowth of clonal subpopulations.
INTRODUCTION

Recent high-throughput sequencing and genomic hybridization studies have revealed substantial intratumoral heterogeneity in cancer patients (1, 2). Sections of single biopsies as well as biopsies taken from primary and metastatic regions revealed a highly complex nonlinear branching clonal evolutionary model as the basis for cancer progression and intratumoral diversity (3–11). Matched diagnostic and relapsed patient samples have also revealed the dynamic nature of a heterogeneous tumor, with the dominant subpopulation at relapse often originating from a minor pre-existing subclone (12–16). As such, spatial and temporal intratumoral heterogeneity presents a fundamental challenge for the rational design of combination chemotherapeutic regimens, which remain the primary treatment for most systemic malignancies.

There have been attempts to study heterogeneity and optimal therapeutic strategies. Many have been theoretical – on examining drug scheduling with generic single drugs or drug combinations on a heterogeneous population containing a sensitive and a resistant subpopulation (17–20). Some of these scheduling strategies have been followed by experimental validations in vitro (21, 22). However, a rational approach to design drug combinations to minimize the effects of intratumoral heterogeneity in a tractable model with experimental validation in vitro and in vivo in preclinical models has been lacking.

Here we apply a computational optimization algorithm constructed on the experimental foundation of known single-drug efficacies for genetically variant cell subpopulations to predict how drug combinations will affect heterogeneous tumors. The most crucial assumption inherent in our algorithm has been demonstrated in our recent work: that commonly used combinations of chemotherapeutics act as linear averages of each component drug against homogenous tumors (23). For example, a two drug combination of drug A and drug B creates a combination selective pressure that resembles a simple weighted sum of A + B. Integration of this experimental foundation with our mathematical framework offers an advance in predictive understanding of how drug combinations influence the fate of heterogeneous tumors, which we successfully validated both in vitro and in vivo.

RESULTS

RNAi-based approach to model heterogeneity and drug combination optimization

Our experimental system derives from the conceptual premise that tumors undergo branched clonal evolution with deregulated oncogene expression and/or tumor suppressor loss providing a basis for transformation, and additional genetic changes accumulating during tumor progression. These additional mutations underlie the development of a heterogeneous tumor population (Fig. 1A, as a simplified
example). An attractive approach to model this heterogeneity in a tractable system amenable to systematic study is the use of an RNAi-based approach (Fig. 1B). Knockdown of specific genes of interest can approximate the loss of function that occurs in tumors, with the combination of multiple shRNA-expressing subpopulations modeling the diversity present in a heterogeneous tumor. We have previously used an RNAi-based approach to elucidate mechanisms of single and combination drug action (23, 24) using Eμ-myc; p19Arf−/− lymphoma cells. Given this dataset of known therapeutic effects of cytotoxic and targeted therapies (Table S1) on individual shRNA-expressing subpopulations, we hypothesized that we could computationally predict superior versus inferior treatment strategies for minimizing subpopulations within a heterogeneous tumor (Fig. 1C).

Our mathematical algorithm is based on integer programming, which identifies the set of drugs that should be present in order to accomplish an aspired goal, or 'objective function'. In the first manifestation here, the defined goal is to minimize outgrowth of specific tumor subpopulations within a known heterogeneous population (see Material and Methods and Supplementary Text for mathematical formulations). The rationale for this goal as a first test of our hypothesis is that current clinical approaches tend to focus on targeting the predominant tumor subpopulation, and the outgrowth of originally minor subpopulations representing resistance to that treatment is at present a daunting clinical problem.

We used this algorithm, in combination with a previously published data set on drug-genotype interactions (23, 24), to identify the most-effective two-drug combinations sampled from a large number of three-component heterogeneous tumor populations (tumors that are comprised of two RNAi-produced sub-populations plus the parental sub-population) (Fig. S1). We then systematically simulated all possible combinations of three-component population and determined the most and least effective two-drug combinations (Table S2). In many cases, the optimal therapy was the same independent of whether we examined the entire heterogeneous population or just the predominate subpopulation. However, there was also a subset of population compositions where the solutions differ (Fig. S2). This suggests that for some heterogeneous tumor populations, we cannot derive the optimal therapy based on solely the predominant subpopulation alone. One such example was a three-component population consisting of the parental Eμ-Myc; p19Arf−/− lymphoma (no shRNA) and subpopulations expressing either a Chk2 (a DNA damage checkpoint regulator) or a Bok (a Bcl2-family cell death mediator) shRNA. The optimal treatment for this tumor was a vincristine (Vin) plus SAHA combination, whereas the worst was an irinotecan (IRT) plus chlorambucil (CBL) combination. Interestingly, if only shChk2 or shBok populations were examined alone, the predicted optimal combination was neither Vin/SAHA nor
IRT/CBL (Fig. 1D). Furthermore, SAHA was not the best single-agent for either shChk2 or shBok alone, but becomes part of an optimal drug combination in a heterogeneous population containing both shChk2 and shBok. This was also the case in examining lymphomas containing an alternative population composition of parental/shChk2/shBim. Thus, consideration of a heterogeneous tumor in its entirety can result in nonintuitive optimal drug combinations, sometimes containing drugs that are not the best single-agent for any subpopulations.

In vitro validation of drug combination on heterogeneous tumor

To experimentally validate these predictions, we used an in vitro fluorescence-based competition assay (Fig. 2A-B and Fig. S3), in which a mixture of parental lymphoma cells and shRNA-expressing subpopulations were exposed to combinations of drugs at controlled doses. Specifically, the drug combinations were dosed such that each drug in the combination contributed equally to a cumulative LD80-90 combination cell killing. GFP- or Tomato-labeled subpopulations enabled us to track the enrichment or depletion of individual subpopulations in the population mixture. Here, we observed that Vin/SAHA effectively maximized the therapeutic response in shBok-containing tumor cells (i.e. enhanced the depletion of shBok-infected cells) while minimizing the selective outgrowth of populations of tumor cells expressing shChk2 (Fig. 2C, Fig. S4). In contrast, IRT/CBL strongly selected for the resistant shChk2 subpopulation, and the parental subpopulation remained sensitive to both combinations. Combination treatment with actinomycin D (ActD)/Erlotinib or Vin/CBL, which were predicted to be optimal treatments if shChk2 or shBok subpopulations were considered individually, were less effective than Vin/SAHA. These predictions were also validated in a different tumor with a distinct population composition (Fig. 2D, Fig. S4). Taken together, these data suggest that combination therapies can be tailored to effectively minimize the effects of a heterogeneous tumor population given some knowledge of the composite subpopulations and their responses to single drugs.

As tumors continuously evolve with the acquisition of new mutations, forming more complex hierarchical structures, we wondered whether we could extend our simple model to account for another layer of complexity. Here, we performed an additional knockdown in our existing subpopulations to form a heterogeneous population consisting of parental, shChk2 (alone), shBok (alone), and shChk2 plus shBok. We then examined the response of these cells to the same four combination therapies and observed that the results were consistent with our predictions, with the optimal therapy still Vin/SAHA (Fig. 2E and F, Fig. S4). We can also apply principal component analysis as a method for dimensionality reduction and visualization on the resulting population, as we have done previously for single populations (25, 26). We observed that the optimal drug combination effectively impacts the population composition towards the
goal of eliminating the shRNA subpopulations (Fig. S5). Thus, for this limited set of tested cases, we have demonstrated successful prediction of effective therapeutic strategy design, by computational integration of the individual drug effects across known target subpopulations, which can produce non-intuitive outcomes.

In vivo validation of drug combination on heterogeneous tumor in preclinical lymphoma model

We next sought to validate our predictions in vivo. The murine Eμ-Myc lymphoma model (27, 28) allows us to perform ex vivo transduction of tumor cells followed by transplantation into syngeneic immunocompetent recipient mice (Fig. 3A). Lymph node, thymus, and spleen are among several sites of tumor dissemination in both donor and recipient tumor-bearing mice. Using this model, we observed substantial intratumoral heterogeneity by ex vivo whole mouse fluorescence imaging (Fig. 3B and Fig. S6). However, when we then analyzed individual lymph nodes, thymus, and spleen of untreated mice using flow cytometry, we observed heterogeneity at the level of individual tumors (Fig. S7). To validate the therapeutic effects in vivo, we determined the optimal dose for each of the individual drugs to ensure there was comparable therapeutic effect on parental tumors from each component drug in the combination (Fig. S8). Using these doses, combination treatment with Vin/SAHA successfully minimized the emergence of any tumor subpopulation, while IRT/CBL enriched significantly for the intrinsic resistant shChk2 subpopulation (Fig. 3C-D). Since our combination therapies were optimized to minimize specific tumor subpopulations relative to the parental lymphoma cells, we also examined the tumor-free survival of mice with heterogeneous lymphoma tumor normalized to that of mice with homogeneous lymphoma tumor. Upon treatments, Vin/SAHA improved the relative tumor-free survival compared to IRT/CBL (Fig. 3E). Taken together with our in vitro results, these data suggest that we can apply a mathematical optimization approach to predict optimal therapeutic strategies for minimizing the emergence of genetically defined tumor component populations and impact the tumor-free survival of mice when presented with a heterogeneous instead of homogeneous tumor.

We further explored the comparison between Vin/SAHA and IRT/CBL on relative tumor-free survival in shChk2/shBok/parental tumors across wide ranges of subpopulation proportions. With a three-component population, all tumor compositions can be represented on a ternary plot, with each corner corresponding to a homogeneous population of a single component and any point within the plot corresponding to some specific composition mixture of the three components. We experimentally determined the tumor-free survival of mice with homogeneous tumor containing one of the three subpopulations for treatment with either of the combination regimens (Fig. S9 and S10). Using this information, we generated a descriptive model showing the efficacy difference between Vin/SAHA and IRT/CBL across all possible...
subpopulation proportions. We observed that introducing heterogeneity (shChk2 and/or shBok subpopulations) into the parental tumor, Vin/SAHA increasingly improves the relative tumor-free survival (Fig. 4A). We can also derive the comparison in terms of absolute tumor-free survival, to inform us of tumor compositions for which Vin/SAHA dominates over IRT/CBL (Fig. S11). In both cases, an example heterogeneous tumor composition (used for the competition assays above), which was not used in generating the model approximates well with the prediction described by the model.

Since our optimization model predicts the effect of drug combinations on the evolution of tumor composition, and our descriptive model approximates the efficacy difference between treatments at specified tumor composition, we can combine these models to examine drug efficacy at multiple time points upon single or consecutive drug treatments. While both drug combinations have minimal effects on changes of tumor composition with empty vector controls (Fig. 4B), the drug treatments predictably affected the heterogeneous tumor (Fig. 4C) – landing at tumor compositions at relapse for which we can also now estimate the therapeutic efficacy comparisons for the next round of treatment. As pre-existing subclones prior to treatment may exist at extremely low frequencies in clinically observed cases, we also examined theoretically the dynamics of tumor heterogeneity with multiple rounds of combination therapy at different initial tumor compositions, including extremes such as 0.1% pre-existing shChk2 and shBok. Sequential IRT/CBL strongly selected for the shChk2 subpopulation whereas Vin/SAHA had less selective pressure on the subpopulations (Fig. S12). Thus, based solely on tumor heterogeneity dynamics (assuming no new induced resistance mechanisms), Vin/SAHA remains as a superior effective drug combination whereas after several rounds of IRT/CBL, the effectiveness of this drug can be dramatically reduced due to the outgrowth of shChk2 in becoming the dominant subpopulation.

Taken together, knowledge of the in vitro efficacy of single drugs on individual subpopulations has allowed us to optimize for a drug combination that minimizes the effects of heterogeneity and improves relative tumor-free survival. We combined additional in vivo efficacy data of the chosen drug combinations on individual subpopulations to approximate the in vivo therapeutic responses over wide ranges of tumor subpopulation proportions in a heterogeneous tumor. These two models enable the tracking of tumor trajectories upon single and multi-course treatments and examination of the effectiveness of subsequent treatment strategies at relapse(s).

DISCUSSION
Starting with Peter Nowell's seminal 1976 paper on the clonal evolution of cancer (29), we have gradually gained the ability to monitor and deconvolute the step-wise alterations in cancer evolution that
underlie intratumoral heterogeneity. Not only are distinct tumor sub-clones found to coexist within the same tumor regions (30, 31), but analyses of biopsies in primary and metastatic tumor sites further suggests that metastatic sub-clones originate from a non-metastatic parental clone in the primary tumor (3-11). Additional changes at the post-transcriptional and epigenetic level can potentially further diversify a tumor population with functional variations (32, 33). This heterogeneous tumor population is also dynamic, as has been shown in the responses to standard combination chemotherapeutic regimens, with pre-existing minor sub-clones expanding to dominate at relapse (12-16). As such, current combination regimens can have unpredictable and/or unintended consequences on the resulting tumor diversity. The original rationale for the use and choice of combination therapies has been primarily to increase the effective and tolerable drug dose, while minimizing resistance (34-38). In theory, this leads to more cancer cell killing and decreases the likelihood that a resistant clone will compromise therapeutic success. However, in light of recent studies showing the extent of intratumoral variation and its clinical implications, it is paramount to incorporate tumor diversity and the expected evolutionary trajectories into rational drug combination design to achieve predictable tumor response, reduce chances of relapse, and prolong patient survival. Our joint theoretical and experimental approach suggests that we can derive drug combinations that can minimize outgrowth of specific subpopulations in a given heterogeneous tumor while enhancing tumor-free survival in mice — provided we know some knowledge of the tumor composition and the response of component subpopulations to single drugs.

Here we used the *Eμ-Myc* lymphoma mouse model as the basis for our computational modeling and experimental validations. The presence of *p19\(^{TR}\)* in this model may remove selective pressure on specific mutations that arise during tumor progression and in response to treatment. Thus, the broad applicability of this approach to other cancer models remains to be seen. Nevertheless, the *Eμ-Myc* lymphoma model is an extremely well-characterized preclinical mouse model and thus a good initial system to tractably address the rational design of drug regimens in the context of genetically diverse tumor. Furthermore, *Eμ-myc* models a defining translocation observed in Burkitt’s lymphoma. Previous targeted molecular analyses on different sites of the same patient and matched sample at diagnosis and relapse revealed shared *Myc* breakpoints and *Bcl-6* mutations, but differential *p53* and *Myc* mutations across the different samples, suggesting clonal evolution spanning across different tumor sites and in response to treatment (39).

The complexity of heterogeneous tumors found in patients undoubtedly complicates applications of this approach. However, our studies highlight basic principles towards rational drug combination design. For example, as was the case for Vin/SASA in this tumor model, a therapeutic strategy may involve finding a
set of drugs that, when combined, maximizes the therapeutic synthetic lethality of certain subpopulation(s), while minimizing the selective pressure for outgrowth of cells bearing undesirable alteration(s). Here, the maintenance of tumor heterogeneity – representing the persistence of a “naïve” pre-treatment state – may be preferable to the undesirable outgrowth of specific subpopulation(s). On the other hand, if we can potentially drive towards the enrichment of therapeutic sensitive subpopulations while preserving overall tumor sensitivity, a sequential treatment approach may be successful as part of a multi-course regimen. Thus, the ability to apply optimization approaches to control the trajectories of tumor composition offer opportunities to maximize the effects of concurrent or successive drug combination regimens.

MATERIALS AND METHODS

Cell culture and chemicals
Murine Eμ-myc; p19arf B-cell lymphomas were cultured in B-cell medium (45% DMEM, 45% IMDM, 10% FBS, supplemented with 2mM L-glutamine and 5µM β-mercaptoethanol). The cell line was tested and shown to be free of mycoplasma using both PCR-based (ATCC) and biochemical-based (Lonza) methods. All drugs were obtained from LC Laboratories, Sigma-Aldrich, Calbiochem, or Tocris Biosciences. For in vivo studies, irinotecan, chlorambucil, and vorinostat were dissolved in DMSO as stock and further diluted in 0.9% NaCl solution prior to injection. Vincristine was dissolved in 0.9% NaCl solution.

shRNA constructs
All shRNAs were expressed in either MSCV-LTR-MIR30-SV40-GFP (MLS) (40) or MSCV-LTR-MIR30-PGK-Tomato (MLT) retroviral vector and were previously validated for knockdown (24). Multiple shRNAs targeting each gene were used for single drug responses to rule out off-target effects (24). Transfection and transduction were performed as previously described (23).

In vitro competition assay
Single and combination drug treatments in vitro were performed as previously described (23, 24) (see Fig. S3). Briefly, lymphoma cells were infected to ~20% of the total population with the indicated retroviruses and were dosed with single or two drug combinations at concentrations needed to achieve 80-90% cell death (LD80-90) on the empty vector control (MLS or MLT) cells (assessed at 48 h). For combination treatments, each drug was dosed to ensure equal contribution of overall killing in the combination. Cell death was assessed with propidium iodide incorporation. Every 24 h, untreated and treated cells were diluted 1:4 and 1:2, respectively. At 72 h, the percentage of GFP+ and/or Tomato+ cells were analyzed on
a cell analyzer FACSScan, FACSCalibur, or LSRFortessa (BD Biosciences). To assess enrichment/depletion of the subpopulations, a log transformed Resistance Index (RI) \( (23, 24) \) was calculated: 

\[
\log_2\text{RI} = \log_2\left[\frac{(L_t - L_{t0})}{(L_u - L_{u0})}\right],
\]

where \( L_t \) and \( L_u \) represents percentage of labeled cells (GFP+ and/or Tomato+) out of the total live population with and without treatment, respectively. See Supplemental Text for derivations of log2RI. Experiments were performed with indicated shRNAs in both MLS and MLT vectors to rule out any vector-specific effects.

**In vivo competition and survival assay**

Two million infected \( E\mu\)-myc; \( p19^{wt}\) lymphoma cells were tail-vein injected into 7-9 weeks female syngeneic recipient C56/BL6 mice. Upon palpable tumor (~12-13 days after injection), mice were treated with single or combination drug regimens. Dosing for each drug was determined to ensure comparable efficacy on control uninfected lymphoma cells. Irinotecan (IRT) was administered at 120 mg/kg i.p. once on day 1; vincristine (Vin) at 1 mg/kg i.p. once on day 1; chlorambucil (CBL) at 10 mg/kg i.p. once on day 1; and vorinostat (SAHA) at 300 mg/kg i.p. on days 1 and 3 and 150 mg/kg on days 2 and 4. Vehicle control consisted of maximum equivalent doses of DMSO and 0.9% NaCl solution. For competition assay, lymphoma cells from thymus, lymph node, and spleen were harvested at sizable palpable tumor after relapse and analyzed on a cell analyzer. For survival studies, tumor-free survival was monitored by daily palpation following treatment.

**In vivo fluorescence imaging**

Fluorescence imaging of mice was acquired using the IVIS imaging system and Living Image software v4.3.1 (PerkinElmer). Two excitations (535nM and 465nM) with 9 emission filters (for a pairwise combination of ten) were used. GFP, Tomato, and background signals were spectral unmixed with resulting composite image further enhanced with brightness and contrast.

**Mathematical optimization and models**

The optimization of combination therapies for minimizing heterogeneous subpopulations was mathematically formulated as a binary integer programming problem. All analyses were performed in MATLAB R2012b (MathWorks). Binary integer programming was solved using the bintprog function in the MATLAB Optimization Toolbox. In determining cases for experimental validation, we sorted the simulation results based on difference in treatment effectiveness (evaluated by log2RI) between therapy optimized based on consideration of entire heterogeneity and that of just the predominant subpopulation. We validated cases with large efficacy differences, including non-intuitive cases (e.g. optimal drug combinations for heterogeneous that does not contain component drugs best for predominant
subpopulation) and cases with different heterogeneous tumor compositions. Refer to Supplemental Text for more details on the mathematical model for optimization and descriptive model on the efficacy comparison between Vin/SAHA and IRT/CBL.

**Rotated principal component analysis**

Rotated principal component analysis was used to visualize the impact of different drug combinations on tumor composition. The data consists of a single matrix with each row representing a different treatment condition (or a pure subpopulation) and each column representing an averaged \( \log_2 \text{RI} \) for each drug. The averaged \( \log_2 \text{RI} \) was derived from matrix multiplication of the enrichment/depletion dataset matrix, \( R \) (see mathematical models section in Supplemental Text) by population composition vector, \( p \), determined experimentally following treatment. Thus, the derived data represents a pharmacological profile for each resulting population following a specific treatment condition. The data was mean centered and unit variance scaled. Principal component analysis (PCA) was performed using SIMCA-P v11.5 (Umetrics). For ease of interpretation, a linear transformation of the principal component space was performed using MATLAB by rotating the parental population projection completely into the second principal component.

**Statistical analyses**

Statistical analyses were performed using Prism v5 (GraphPad). Two-tailed Student’s t-tests and one-way ANOVA (with Bonferroni post-hoc test) were used, as indicated. Error bars represent mean ± s.e.m. Logrank (Mantel-Cox) test was used for comparison of survival curves.

**ACKNOWLEDGEMENTS**

We thank Luke Gilbert and Peter Bruno for advice on experimental procedures; members of the Lauffenburger and Hemann lab for helpful discussions; and Peter Bruno, Eleanor Cameron, Miles Miller, and Aaron Meyer for comments on the manuscript. Funding was provided by Integrative Cancer Biology Program Grant U54-CA112967-06 (to M.T.H. and D.A.L.). B.Z. is supported by the NIH/NIGMS Interdepartmental Biotechnology Training Program 5T32GM008334. This work was also supported in part by the Koch Institute Support (core) Grant P30-CA14051 from the National Cancer Institute.

**AUTHOR CONTRIBUTIONS**

B.Z., J.R.P., D.A.L. and M.T.H. designed research. B.Z. performed experiments and mathematical calculations. B.Z., J.R.P., D.A.L. and M.T.H. analyzed data. B. Z., D.A.L. and M.T.H. wrote the manuscript.
REFERENCES

Fig. 1. A strategy for modeling intratumoral heterogeneity and mathematical optimization. (A) A simplified schematic of tumor evolution, with deregulated Myc expression and \( p19^{Arf} \) loss, followed by additional loss-of-function mutations. (B) RNAi-based modeling of intratumoral heterogeneity. Each shRNA knockdown models a specific loss-of-function event. Mixture of these subpopulations creates a heterogeneous tumor population. (C) A schematic of how mathematical optimization can be applied to drug design for tumor heterogeneity. We have previously acquired a dataset on the response of specific shRNA knockdowns to a set of single chemotherapeutic and targeted agents (23, 24). Here, using this dataset and given a particular population composition, we applied an optimization approach (see Material and Methods) to determine drug combinations that are best and worst at treating all subpopulations. (D) Two top “hits” (i.e. population compositions) derived from computational simulation demonstrating that the optimal drug combination predicted is different depending on whether we examine the entire heterogeneous tumor population or only a particular subpopulation (see Fig. S2 and Table S1 for summary of all simulation results).
Fig. 2. *In vitro* validation of predicted effects of combination therapies on subpopulation composition. (A) A schematic of the *in vitro* competition assay (see also Fig. S3). *Eμ-Myc; p19<sup>trf</sup>*/<sup>-</sup> lymphoma cells were retrovirally transduced with the desired shRNA(s). Mixed populations were treated with combination treatment and analyzed at 48 h. (B) A representative flow cytometry result of no treatment and the indicated combination treatments of a mixed population of tumor cells containing parental, shChk2 (Tomato-labeled), and shBok (GFP-labeled). (C-E) *In vitro* competition assay results for different population composition using different combination treatments. Drug combinations predictably enriched/depleted subpopulations. (F) Correlation between predictions made from mathematical model and the actual experimental results. Gray data points represent individual subpopulations (e.g. shChk2 only) and blue data points represent combined subpopulations (e.g. shChk2 and shBok). Data shown are
mean ± s.e.m. of three independent experiments, **$P<0.01$, ***$P<0.001$ (ANOVA with Bonferroni post-hoc test).
Fig. 3. *In vivo* validation of effects of combination therapies on subpopulation composition. (A) A schematic of the *in vivo* competition assay. *Eμ-Myc; p19ΔRt/c* lymphoma cells were retrovirally transduced with desired shRNA, and mixed populations of tumor cells were tail-vein injected into recipient mice. Combination treatments were given at presentation of palpable tumor. Tumor cells were collected and analyzed at relapse. (B) Representative fluorescence imaging of vehicle-treated mice with mixed populations of parental, shChk2 (GFP-labeled), and shBok (Tomato-labeled) infected tumor cells, showing intratumoral heterogeneity *in vivo*. (C) Representative *in vivo* competition assay flow cytometry analysis of relapsed tumors following *in vivo* treatment with the indicated combination therapies. (D) An *in vivo* competition assay showing the enrichment or depletion of subpopulations of shRNA infected tumor cells. Vin/SAHA and IRT/CBL treatment data for the combined shBok and shChk2 populations are shown in lymph node, thymus, and spleen. Drug combination predictably enriched/depleted subpopulations, in agreement with *in vitro* results (Fig. 2). Data shown are mean ± s.e.m of three independent experiments (with 4-5 mice per experiment). **P<0.01 (two-tailed Student's t-test). (E) Relative tumor-free survival for each combination treatment on mice transplanted with heterogeneous parental/shChk2/shBok tumor, with days normalized to the median tumor-free survival of treated mice with homogeneous parental tumor. Vin/SAHA improved relative tumor-free survival when compared to IRT/CBL. Data were compiled from four independent experiments. *P* value was calculated using a log-rank test.
Fig. 4. Vin/SAHA is superior to IRT/CBL in extending relative tumor-free survival in a three-component heterogeneous tumor across varying subpopulation proportions. (A) A ternary plot showing the comparison between Vin/SAHA and IRT/CBL in terms of relative tumor-free survival in a shChk2/shBok/parental tumor (i.e. tumor-free survival of heterogeneous tumor normalized to that of a parental-only empty vector control) at varying subpopulation proportions. Each white dot represents a tumor composition for which experiments were performed to determine tumor-free survival in mice. Three out of the four dots at the corner of the ternary plot represents the homogeneous tumor that was used to generate the model. The position of the internal white dot approximates the heterogeneous tumor composition at the end of the experiment in vehicle-treated mice. (B-C) Predicted and actual experimental trajectories of tumor with empty vector control parental/mls/mlt (B) or heterogeneous tumor with parental/shChk2/shBok (C) upon treatment with Vin/SAHA or IRT/CBL in vivo. For heterogeneous tumor (C), the ternary plot was overlay with heat map in (A) showing the therapeutic efficacy comparisons between Vin/SAHA and IRT/CBL. Dotted circle denotes initial tumor composition; solid circle denotes the mean final tumor composition (of pooled lymph nodes per mouse) at relapse following treatment. Predicted trajectories are shown with gray circles and dotted arrows; experimental results of trajectories are shown in white circles and solid arrows.
**Supplementary Figure S1.** Monte Carlo sampling results for two-drug combinations on a thousand three-component tumor populations. Drug combination was optimized using an integer programming algorithm with the mathematical model described in Supplemental Text. Briefly, the optimization finds the optimal two-drug combination that minimizes the overall tumor heterogeneity, provided knowledge of the tumor composition and effects of single drugs on single subpopulations. The statistical sampling approach enables an examination of the frequency of a drug in optimal drug combinations, a characteristic that is independent on knowledge of an exact instance of tumor composition. (A) A uniform distribution of the subpopulations can be used for Monte Carlo sampling. (B) The resulting distribution of drugs in the two-drug combination optimized based on sampling from a subpopulation distribution in (A). (C) An alternative distribution of subpopulations, based on frequency of mutations observed in hematopoietic cancers (using COSMIC v67). (D) The resulting distribution of drugs in two-drug combination based on sampling from a subpopulation distribution in (C).
Supplementary Figure S2. Simulation results for two-drug combination on three-component tumor populations (all possible tumor compositions were enumerated with subpopulation proportions discretized in 1% increments). Histogram shows the distribution of the difference in predicted efficacy (reported here as Δlog2RI) on each heterogeneous tumor between two two-drug combinations: one optimized based on considering the entire heterogeneous tumor versus one optimized based on considering just the heterogeneous tumor’s predominant subpopulation.
Supplementary Figure S3. A schematic overview of the in vitro competition assay for single and combination drug treatments on individual shRNA population variants in heterogeneous lymphoma population. Parental Eμ- myc; p19Arf−/− lymphoma cells were infected with indicated retroviruses. For heterogeneous mixed population, the infected cells were sorted and mixed at indicated ratios prior to treatment. Single drug or drug combinations were dosed at concentrations to achieve a cumulative 80-90% cell death. Enrichment or depletion of the subpopulations was assessed using flow cytometry at 48 h.
Supplementary Figure S4. Bar graphs showing in vitro competition results for parental/shChk2/shBok,
parental/shChk2/shBim and parental/shChk2/shBok/shChk2 plus shBok populations treated with four different drug combinations. Plots show enrichment/depletion of individual subpopulations. Each shRNA was tested in both Tomato-labeled MLT and GFP-labeled MLS vector to rule out vector-specific effects. No significant enrichment/depletion was observed with the MLS and MLT empty vectors alone.
Supplementary Figure S5. Rotated principal component analysis (PCA) of in vitro validation results. Optimal treatment moves the initial heterogeneous tumor state towards the objective – the minimization of clonal outgrowth.
Supplementary Figure S6. Representative Ex vivo whole mouse fluorescence imaging of vehicle-treated mice with mixed populations of parental/shChk2 (GFP-labeled)/shBok (Tomato-labeled) or empty vector control parental/mls (GFP-labeled)/mlt(Tomato-labeled) tumor cells. Mice were imaged 6 days following the emergence of palpable tumors. Mouse 4 with parental/shChk2/shBok was used for Fig. 3B.
**Supplementary Figure S7.** *Ex vivo* whole mouse fluorescence imaging followed by flow cytometry of individual tumors from mice with mixed populations of parental/shChk2 (GFP-labeled)/shBok (Tomato-labeled) or empty vector control parental/mls (GFP-labeled)/mlt (Tomato-labeled). Individual tumors were analyzed by flow cytometry and the adjacent pie charts show the proportion of the three tumor subpopulations.
Supplementary Figure S8. Determination of the optimal dose for individual drugs to ensure comparable therapeutic effect *in vivo* on control uninfected *Eμ-myc; p19* tumor from each component drug in the combination. (A) A diagram showing the transplanatation and treatment regimen. A total of 2 million uninfected cells were tail-vein injected into recipient mice. Mice were treated upon palpable tumor. (B) A list of optimized doses used for determining tumor-free survival. (C) A Kaplan-Meier curve showing tumor-free survival of mice treated with the doses indicated in B.
Supplementary Figure S9. Kaplan-Meier curves showing tumor-free survival of mice transplanted with (A) shChk2, (B) shBok or (C) parental homogeneous tumors, or (D-F) heterogeneous parental/shChk2/shBok tumors and treated with Vin/SAHA, IRT/CBL, or vehicle control. P value was calculated using a log-rank test.
Supplementary Figure S10. Kaplan-Meier curves showing tumor-free survival of mice transplanted with (A) shChk2, (B) shBok or (C) parental homogeneous tumors treated with Vin/SAHA or IRT/CBL. The data were derived from that in Supplementary Fig. S9, but now with days normalized to the median tumor-free survival of mice transplanted with homogeneous parental *Eμ-myc; p19*ΔΔ/*tumors. *P* value was calculated using a log-rank test.
**Supplementary Figure S11.** A ternary plot, generated using a descriptive model (see Supplementary Text), showing the comparison between Vin/SAHA and IRT/CBL in terms of absolute tumor-free survival in a shChk2/shBok/parental tumor at varying subpopulation proportions. Each white dot represents a tumor composition for which experiments were performed to determine the tumor-free survival of mice. Three of the dots are located at the corner of the ternary plot and represent homogeneous tumors that were used to generate the model. The position of the internal white dot approximates the heterogeneous tumor composition at the end of experiment in vehicle-treated mice.
Supplementary Figure S12. Theoretical tumor composition trajectories following multiple rounds of combination treatment with (A) Vin/SAHA or (B) IRT/CBL. Several initial tumor compositions, denoted with dotted circle in different colors, were selected as initials points for this model. Three rounds of treatments were simulated, with the tumor composition following each round of treatment denoted in a solid circle, and in increasing color intensity for each successive round (see Supplemental Text for details
of mathematical model). Vin/SAHA had less selective pressure for outgrowth of subpopulations, whereas IRT/CBL strongly selected for shChk2, leading to its outgrowth after only a few rounds of treatment.
Chapter III

Intratumor heterogeneity alters most effective drugs in designed combinations


Supplemental Text and Tables are available online.
ABSTRACT
The substantial spatial and temporal heterogeneity observed in patient tumors poses considerable challenges for design of effective drug combinations with predictable outcomes. Currently, the implications of tissue heterogeneity and sampling bias during diagnosis are unclear for selection and subsequent performance of potential combination therapies. Here, we apply a multi-objective computational optimization approach integrated with empirical information on efficacy and toxicity for individual drugs with respect to a spectrum of genetic perturbations, enabling derivation of optimal drug combinations for heterogeneous tumors comprising distributions of subpopulations possessing these perturbations. Analysis across probabilistic samplings from the spectrum of various possible distributions reveals that the most beneficial (considering together efficacy and toxicity) set of drugs changes as the complexity of genetic heterogeneity increases. Importantly, significant likelihood arises that a drug selected as the most beneficial single-agent with respect to the predominant subpopulation in fact does not reside within the most broadly useful drug combinations for heterogeneous tumors. The underlying explanation appears to be that heterogeneity essentially homogenizes the benefit of drug combinations, reducing special advantage of a particular drug on a specific subpopulation. Thus, this study underscores the importance of considering heterogeneity in choice of drug combinations, and offers a principled approach toward designing the most likely beneficial set even when the subpopulation distribution is not precisely known.

SIGNIFICANCE STATEMENT
Tumors within each cancer patient have been found to be extensively heterogeneous both spatially across distinct regions and temporally in response to treatment. This poses challenges for prognostic/diagnostic biomarker identification and rational design of optimal drug combinations to minimize reoccurrence. Here we present a computational approach incorporating drug efficacy and drug side effects to derive effective drug combinations and study how tumor heterogeneity affects drug selection. We find that considering subpopulations beyond just the predominant subpopulation in a heterogeneous tumor can result in non-intuitive drug combinations. Additional analyses reveal general properties of effective drugs. This study highlights the importance of optimizing drug combinations in the context of intratumor heterogeneity and offers principled approach toward their rational design.
INTRODUCTION

Genetic intratumor heterogeneity has long been appreciated as present in cancer patients (1). Recent sequencing studies have further revealed the extent of this tumor diversity, arising from highly complex clonal evolutionary processes (2). This phenomenon has been observed in many solid (3–5) and hematopoietic cancers (6–8). Moreover, treatments can also have dramatic effects on the tumor composition – with preexisting subclone at diagnosis often attain domination at relapse (9–12). To meet this challenge, rational drug combination treatments need to be designed in a manner accounting for intratumor heterogeneity to enable reoccurrence to be overcome more predictively.

A long history exists of theoretical studies aimed at drug optimization in cancer therapy. Some have used differential equation models formulated as deterministic optimal control problems (13–16), others stochastic birth-death process models (17–21), to examine treatment regimens on tumors comprising a sensitive population along with a few resistant subpopulations. However, these studies – many of which deal with only generic scenarios – have primarily focused on drug scheduling, investigating the effects of frequency and dose intensity of drugs on resistance potential. Practical applications of such results have been limited, albeit some of these strategies have been more recently combined with experimental validation to examine single-drug scheduling in vitro and in xenograft models (22, 23).

The extensive number of chemotherapy and targeted therapeutics in clinical and preclinical use presents a large search space to determine even the choice of component drugs in a drug combination for best treatment of a heterogeneous tumor. Perhaps the closest endeavor has been a theoretical analysis of a minimal set of drugs that maximize the coverage of molecular target variants, with respect to structural considerations (24). Nonetheless, to our knowledge there are no previous systematic studies of combination drug design examining efficacy and toxicity in the context of tumor heterogeneity, aimed at the questions: [a] how heterogeneity affects the utility of drug combinations; and [b] how current clinical practice of tumor diagnosis, which biases toward focus on the predominant subpopulation, affects design of drug combinations (Fig. 1A).

To address these questions, we employ here a computational optimization approach in concert with previously reported experimental data across a range of specific drugs effects on a spectrum of particular RNAi perturbations characterizing genetic variations in model tumors. Our goal is to determine the best drug combinations for together minimizing all subpopulations significantly present in a heterogeneous tumor. We have previously experimentally validated our model and the computational optimization predictions on a two-drug combination with three-component heterogeneous tumors in vitro and in vivo in
the presence of information about the subpopulation proportions (25). Here, we dramatically extended our model, to examine more complex heterogeneous tumor compositions and greater number of component drugs, and most importantly to integrate over samplings from all possible subpopulation distributions via computational simulation. This latter integration facilitates understanding of which drugs should be included in the combination treatment even in absence of quantitative information on the subpopulation distributions for a given tumor, as would be typical in clinical situations. Among insights gained, a key principle is that there can be significant differences between the combinations predicted to be most beneficial for the most predominant single subpopulation and those more beneficial when multiple subpopulations are taken into account. This suggests that gaining at least some degree of information concerning heterogeneity of a primary tumor upon diagnosis, even if not quantitative or complete, can lead to an indicated drug regimen different from what would be viewed as best if only the predominant genetic signature was obtained. Moreover, consistent with our recent experimental findings (26) our simulations demonstrate that drug effects are homogenized by tumor heterogeneity over essentially a full range of conceivable subpopulation distributions. Consequently, the effectiveness of drugs most likely to be found beneficial in combination for heterogeneous tumors can be characterized by average efficacy and robustness across the subpopulations rather than by extreme efficacy on a particular subpopulation. Taken together, our results offer conceptual principles for designing drug combinations in the context of intratumor heterogeneity.

RESULTS

Conceptual and computational approach. Our premise is that a typical tumor comprises multiple diverse subclones along with a dominant subpopulation, and that dynamic changes in the distribution of these populations can be influenced by drug treatments. The currently prevalent clinical practice for diagnosis of tumor type and drug regimen is based on histological and/or biomarker identification and is generally biased toward the predominant subpopulation as a result of its greater representation in the population (2, 27). However, minor subpopulations are important in determining how a tumor responds to treatments and can be responsible for resistance and relapse (28). An immediate critical question is: how should consideration of heterogeneity instead of simply the predominant subpopulation affect drug combination design (Fig. 1A)?

Our approach integrates a theoretical framework with empirical experimental information. The theoretical framework is a multi-objective linear optimization algorithm that simultaneously incorporates efficacy (the desired tumor cell killing) and toxicity (adverse side-effects on patients) for drug combination effects on a large number of heterogeneous tumor compositions (see Supplementary Text for methodological
The calculations involved in the optimization algorithm incorporate effects of each individual drug on each tumor subpopulation, sampling from among conceivable compositions of subpopulations and combinations of drugs (Fig. 1B). These compositions and drugs come from previous experimental information (26, 29). The tumor subpopulations are taken from a battery of shRNA knockdowns of genes in the DNA damage repair and apoptosis pathways in the murine Eμ-myc; p19^{−/−} lymphoma cell line (derived from a well-established pre-clinical Eμ-myc mouse model of human Burkitt’s lymphoma (30, 31)), and the drugs are taken from a series of commonly used chemotherapeutics and targeted therapeutics (Fig. S1). A key finding in this previous work is that overall efficacy of multiple drugs in combinations is typically approximated by linear combinations of the individual drug efficacies when they are applied together at overall LD80-90 concentration. Such additivity allows for use of linear objective function for efficacy with linear constraints in our optimization algorithm.

Of course, design of drug combinations involves constraints and conflicting objectives, raising multiple tradeoffs to consider simultaneously. For instance, tradeoff between efficacy and toxicity prohibits maximizing the number of drugs without exceeding the tolerable toxicity. Previous analyses of optimal control theory-based design of chemotherapeutic regimens have long used toxicity (e.g., in terms of maximal/cumulative drug concentrations) as constraints or secondary objectives in their mathematical formulations (14). Here, our framework instead posits multi-objective optimization to concomitantly maximize overall efficacy and minimize overall toxicity (Fig. 1C and Supplemental Text).

As one of the primary considerations in clinical drug combination regimens is non-overlapping toxicity (again, adverse side effects), we formulated our drug toxicity model as a linear model, with the goal of minimizing overall toxicity with constraints on a maximal allowable toxicity for each side effect (e.g., myelosuppression, gastrointestinal effects, etc.). This formulation effectively captures nonlinear behaviors as a result of non-overlapping drug combinations. Subsequent analyses will first be based on a symmetric toxicity profile – that is each drug has the same toxicity unit of 1. This will later be relaxed to include actual asymmetric toxicity profile of drugs in our efficacy dataset.

Solving a multi-objective optimization problem, concomitantly incorporating efficacy and toxicity, results in a discrete set (known as a ‘Pareto optimal set’) of feasible solutions (i.e., optimal drug combinations) rather than a single “absolute best” solution (Fig. S2). The set is discrete because each solution represents an integer number of drugs included in the combination, but it resides upon a continuous curve termed the ‘Pareto frontier’. Each solution in the Pareto optimal set has the property that an increase in one objective (say, efficacy) results in a decrease at least one or more objectives (then, toxicity). We enumerate the
Pareto optimal solutions via linear programming calculations (see Supplementary Text for technical details).

Effects of tumor heterogeneity on drug combination efficacy and toxicity tradeoffs. As a first manifestation of our analysis, we used our empirical efficacy dataset (26, 29) in concert with a symmetric toxicity profile. In this circumstance, we expect that the algorithm would produce as optimal solutions drug combinations with the most efficacious drug plus additional drugs incorporated into the combination in order of decreasing efficacy. Partnering a symmetric toxicity profile with decreasing efficacy would suggest a deviation from linearity from the Pareto frontier curve as the number of drugs in the combination is increased. Fig. 2A shows a representative example of this anticipated tradeoff between toxicity and efficacy on the Pareto frontier for the homogeneous tumor population (100% shp53). Interestingly, however, as heterogeneity is introduced to the population (55% shp53 plus 13 minor subpopulations), the shape of the Pareto frontier becomes surprisingly linear (Fig. 2B and Table S2). This behavior, of shift from nonlinear Pareto frontier curve to linear as a tumor becomes more heterogeneous, was also observed with other subpopulation distributions (Fig. S4 and Table S3). The meaning of this in tumor treatment terms is that heterogeneity homogenizes the benefit of drug combinations, because each additional drug has differential effects on each subpopulation.

In fact, most clinically used chemotherapeutic regimens consist of multiple drugs, such as ABVD (doxorubicin, vinblastine, bleomycin, dacarbazine), Stanford V (vinblastine, doxorubicin, vincristine, bleomycin, mechlorethamine, etoposide, and prednisone), or BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, prednisone) for Hodgkin’s lymphoma. Accordingly, we imposed an upper limit of six drugs in our subsequent analyses. The consequence of this limit for the same heterogeneous tumor in Fig. 2B is shown in Fig. 2C (the corresponding Pareto frontier) for direct comparison. Fig. 2D enumerates the drugs elicited by the algorithm for each of the regimens, from one-drug to six-drugs.

Monte Carlo analyses of drug combinations in diverse heterogeneous tumor populations. The results above were obtained for single heterogeneous tumors possessing specified subpopulation distribution. In practice, from a clinical tumor biopsy one might gain data on the genetic variants present but not with explicit quantitative distribution proportions. In order to deal with this challenge, we performed Monte Carlo sampling of heterogeneous tumors from among possible quantitative distributions for a given set of genetic variants and analyzed the resulting frequency distribution of component drugs in six-drug combinations. Overall 10,000 heterogeneous populations were generated with each population sampled
based first on the number of subpopulations, followed by the subpopulation genetic variants, and subsequently the subpopulation proportions. Upon drug combination optimization, we obtain predictions of the six-drug combinations most likely to be beneficial for a heterogeneous tumor with the specified genetic variants, averaged over potential unknown quantitative proportions. These genetic variants are characterized by a set of up to 30 shRNA knockdowns, targeting genes in the DNA damage repair and apoptosis pathways (Fig. S1). The component drugs most frequently included in the optimal drug combination for this class of heterogeneous tumor were rapamycin, 5-FU, vincristine, and sunitinib (Fig. 3A); we term these the ‘dominant’ drugs. Although these drugs do not covary based on clustering analysis over all populations (Fig. S5), when we analyzed the frequency of drug inclusion as a function of tumor complexity (i.e., number of subpopulations in the heterogeneous tumor), we observed a strong dependence for select number of drugs – some positively and some negatively correlating with tumor complexity (Fig. 3B, 3C). For instance, with increasing tumor complexity we observed an increase in the frequencies of inclusion of the four drugs noted above but a decrease in the frequencies of some other drugs (such as NSC3852, temozolomide, roscovitine, cisplatin, mitomycin C, and camptothecin). In Fig. 3B these relationships are illustrated with positive or negative trends, respectively, and Fig. 3C quantifies them in terms of Spearman correlation coefficient values. Furthermore, tumor complexity is a crucial characteristic because this strong correlation is abrogated when we control for tumor complexity (by examining the subsequent partial correlation) (Fig. S6).

To further expand generalizability of our insights, we also analyzed combinations of different numbers of drugs other than six; i.e., representing other regimens along a Pareto frontier. The frequency distributions of the component drugs were found to be similar among the various N-drug combinations (annotated as ‘Regimen N’), as calculated in terms of the correlation between the drug frequency distribution in Regimen N versus that in Regimen M (Fig. 3D and Fig. S8). In other words, the drugs more and less dominant in a regimen combining N drugs are similarly more and less dominant in a regimen combining M drugs. Thus, the behavior of drug dominance among a constellation of potential drugs as a function of heterogeneous tumor complexity is consistent across a broad range of multi-drug regimens.

We next asked how the drug frequency distributions might change if we determine the optimal combinations based solely on the predominant tumor subpopulation instead of considering all the subpopulations present in each heterogeneous tumor. Using the same set of 10,000 theoretical tumors, we found only small changes in the overall distribution of drug frequencies within optimal combinations under this approach (Fig. 4A, in comparison to Fig. 3A). However, quite surprisingly, the correlation between drug dominance and tumor complexity is dramatically abrogated (Fig. 4B in comparison to Fig.
3B and Fig. 4C in comparison to Fig. 3C). For example, choice of including drugs such as rapamycin, 5-FU, vincristine, and sunitinib no longer strongly depends on tumor complexity, but instead varies sensitively with the specific individual predominant subpopulation. As a consequence, the drug distributions found in optimal combinations across N-drug regimens may be very different from those for M-drug regimens (Fig. 4D and Fig. S9, in striking contrast to Fig. 3D and Fig. S8). In addition, changes in correlations are minimal when compared before and after controlling for tumor complexity, further suggesting that tumor complexity is no longer an important factor (Fig. S7). Hence, optimal design of combination drugs based on only a predominant subpopulation – i.e., a tumor assumed to be homogeneous – can potentially exclude drugs that work effectively for highly heterogeneous tumors.

We therefore proceeded to analyze in detail the effects of these two alternative optimization perspectives, consideration of the entire heterogeneous tumor versus just the predominant subpopulation – on the drug combinations predicted to be optimal for tumors whose complexity resides within the sampled 10,000 populations. We observed that there was a large proportion of tumors for which the optimal drug combinations were disparate between the two perspectives (Fig. 5A), with the fraction of optimal drug combinations for a given tumor found to be different when taking heterogeneity into account versus not taking it into account increasing from about 40% for 1-drug regimens to almost 70% for 6-drug regimens. Quantitatively, the difference in efficacy between the drug combinations optimized based on the two perspectives followed a roughly exponential profile (Fig. 5B). This means that cases for which the efficacy difference falls towards the tail of the profile will exhibit dramatic disparities in therapeutic outcome when solely the predominant subpopulation is considered for determining the best drug.

Now, it is conceivable that a drug most frequently deemed worthy of inclusion within a regimen optimized for a predominant subpopulation might generally also be found frequently included within a regimen determined from considering the entire tumor complexity. Therefore, we probed this notion again using our 10,000 heterogeneous tumor sample, for each case asking whether the drug combination optimized based on the entire tumor complexity also contains the single-best drug for the predominant subpopulation. Notably, there were a substantial number of regimens for which the drug combination optimized based on the entire heterogeneity does not contain the single-best drug for the predominant subpopulation (Fig. 5C). This proportion expectedly decreased with increasing number of drugs in the combinations; for instance, for 2- and 3-drug regimens 39% and 23% of the cases, respectively, did not include the drug that would have been best for the predominant subpopulation. We obtained similar conclusions when we performed the same analyses on drug combinations optimized based on also the second largest subpopulations (Fig. S10). To confirm the robustness of these conclusions, we performed
these same analyses, but now decomposed the results with respect to the predominant and second largest subpopulation proportions (Fig. S11). We found that the proportion of regimens with different drug combinations as well as those containing single-best drugs for predominant (and second largest) subpopulations is robust near boundary cases across different predominant subpopulation proportions – and only begins to expectedly decrease when the predominant subpopulation becomes very large compared to the remaining subpopulations (Fig. S12). We also performed Monte Carlo sampling with over 13,000 heterogeneous populations near the boundaries, with similar results (Fig. S13). This firmly demonstrates that selecting drug combinations for a heterogeneous tumor may often not follow from simple intuition – which seemingly would involve including the drug best for the predominant subpopulation.

Taken all together, these analyses indicate that intratumor heterogeneity can dramatically influence drug combinations such that the chosen regimen can have strongly disparate outcomes based on whether we examine the entire heterogeneity or just the predominant subpopulation. This insight argues for the necessity of considering at least even qualitative features of heterogeneity in selecting combination regimens.

Sensitivity analysis of Monte Carlo sampling. We next sought to examine in greater detail the specific dependencies of particular drug choice on the tumor subpopulations present. To accomplish this, we performed a global sensitivity analysis by calculating the correlation between a given tumor subpopulation (i.e., shRNA knockdown here) and the categorical outcome of whether a particular drug was chosen by the optimization algorithm to be included in the combination regimen for the 10,000 sampled tumors; this is termed a point biserial correlation. In general there were varying degrees of subpopulation dependency for the different drugs, negative as well as positive (Fig. 6A). In addition, the more frequently chosen drugs exhibit broader distributions of point-biserial correlation values than do the less frequently chosen drugs (Fig. 6B). Interestingly, the less frequently chosen drugs were typically characterized by a very tight range of dependencies and were most distinctly characterized by outliers with respect to particular tumor. In other words, the more frequently chosen drugs are relatively robust to the uncertainty concerning quantitative features of tumor heterogeneity, whereas the less frequently chosen drugs are problematically sensitive to precisely which subpopulations are present and in what quantitative proportions. This insight can be captured by the kurtosis of the distribution of the point-biserial correlation values, which was found to be strongly negatively correlated with the drug’s frequency (Fig. 6C). In addition to robustness, we also observed that, as expected, drug frequency was correlated with mean efficacy averaged over the entire set of 10,000 tumor samples. Taken together, all
these results from sensitivity analysis indicate that the most frequently chosen drugs for optimal drug combinations for heterogeneous tumors of uncertain subpopulation distribution are characterized by their mean and robustness in efficacy.

**Multi-objective optimization comprising particular drug toxicity along with particular drug efficacy.** Our multi-objective optimization model allows for the incorporation of multiple objectives in deriving optimal drug combinations, with whatever sophistication in objective might be desired. As such, we extended our analysis with greater nuance by recognizing that the toxicities differ – that is, are asymmetric – across systemic tissue types for each given drug and include this recognition in our objective functions. Accordingly, we incorporated actual dose-limiting toxicity information for each particular drug in the dataset (Fig. S15A). We observed that now a larger set of choices for drugs could be incorporated for each multi-drug regimen along the Pareto frontier (Fig. S14) and that for myelosuppression and gastrointestinal effects, two of the most common side effects for the drugs in our dataset, the distribution in the number of overlaps between drugs in toxicity was effectively shifted to minimize such overlaps (Fig. S15B, Fig. S15C, and Fig. S16). Hence, a multi-objective optimization approach enables the incorporation of multiple drug characteristics in examining tradeoffs for rational drug combination design in the context of intratumor heterogeneity.

**DISCUSSION**

The substantial spatial and temporal intra-tumor heterogeneity that inevitably exists in cancer patients presents a fundamental challenge for the rational design of effective drug combinations. Here, we applied a multi-objective optimization approach, grounded in empirical experimental data. These data comprised quantitative effects of commonly used chemotherapeutics and targeted therapeutics on subpopulations of murine *Eμ-myc; p19arf−/−* lymphoma cell line generated via shRNA knockdowns of genes in the DNA damage repair and apoptosis pathways. The cell line was derived from the well-established *Eμ-myc* mouse model of human Burkitt’s lymphoma (30), and the existing *p19arf−/−* potentially captures the early events in tumor evolution (31). Our computational analysis on these data generated predictions of optimal drug combinations for tumors in which the genetic heterogeneity is qualitatively characterized (in terms of the shRNA knockdowns) but quantitatively uncertain with respect to the various relative proportions in the tumor. An initial key result is that tumor heterogeneity can effectively homogenize drug efficacy, so that the most effective drug combinations are those that best kill the broadest range of subpopulations. Notably, we found that optimizing drug combinations based on consideration of the entire tumor heterogeneity instead of just the predominant subpopulation can result in non-intuitive optimal drug combinations. As such, knowledge of single “best agent” for each individual subpopulation does not
promise that it will be an optimal choice for included within the overall drug combination when the entire heterogeneous tumor is considered. We have recently demonstrated successful validation of our optimization approach for selected tumor subpopulation distributions in both in vitro cell culture and in vivo mouse contexts (25).

The substantial tumor complexity that exists in patients and the detection limit of diagnostic tools undoubtedly casts uncertainty and incomplete information on the underlying tumor composition for each patient. However, our statistical sampling and sensitivity analyses offer guiding principles for the characteristics of effective drugs under tumor diversity. In particular, we discovered that a certain set of drugs dominated the solutions for optimal drug combination at increasing tumor complexity, and such drugs were associated a higher average efficacy and robustness. This indicates that in the absence of complete knowledge of a tumor composition we can nonetheless still apply de novo design of optimal drug combinations based on drugs’ average efficacy and robustness properties. Ultimately this requires experimental validation in a clinically relevant model. Although our dataset, acquired based on in vitro assays using clinically relevant genes and chemotherapies, is comprehensive on knowledge of efficacy of single agents on single subpopulations, such complete knowledge has yet be realized in the clinical setting for experimental validation. Nevertheless, recent large-scale efforts such as The Cancer Genome Atlas and the Cancer Cell Line Encyclopedia (32) are beginning to bring resolution on patient tumor composition and subpopulation responses. As these types of information accumulate, this offers opportunities to potentially deconvolute and derive efficacy data of single drug on single subpopulations, allowing the incorporation of this methodology to derive drug combinations sampled based on a prior distribution of known genetic variants of a cancer type. This would enable the experimental validation on a large number of clinical samples, with responses tracked longitudinally following treatment.

Additional drug-drug interaction and tumor dynamics considerations undoubtedly add to the complexity of drug combination optimization. These include nonlinear drug-drug efficacy (e.g. drug synergy or antagonism) and toxicity interactions. Specific genetic alterations can also affect subpopulation interactions and mutation rates with consequences on the evolutionary trajectories of tumor, in the absence of drug selection. These additional rate parameters, generally used in stochastic and deterministic differential equation based models of tumor evolutionary dynamics, would be particularly useful to consider in the design of sequential treatment strategies. Moreover, while our model optimized for overall efficacy by targeting all subpopulations, an optimal drug combination will require additional understanding on how differential selective pressures imposed by drugs affect secondary resistance potential. This issue has especially been studied in antibiotic resistance, where strong selection by
synergistic drug combinations may actually increase the risk of resistance (33, 34). Ultimately, just as multi-objective optimization approaches can be applied to derive small molecule structures with desired polypharmacological profiles (35), a multi-objective optimization with considerations of these additional properties provides one potential promising approach to optimize drug combination in the context of tumor heterogeneity.

METHODS
Full details of methods and mathematical models can be found in Supplemental Text. Briefly, efficacy dataset was derived previously using an in vitro competition assay with murine Eμ-myc; p19ARF−/− lymphoma cells partially infected with a specific shRNA and treated with broad range of chemotherapies (26, 29). The toxicity dataset was a binary matrix of known dose-limiting toxicities for each drug. The optimization was formulated as a multi-objective optimization model with the objective of maximizing efficacy while minimizing toxicity. The optimization was solved iteratively with linear programming using Matlab 2012b (Mathworks) and CPLEX 12.5. Monte Carlo sampling was performed with simulation of 10,000 heterogeneous tumors drawn from a specified distribution of subpopulations, followed by drug combination optimization on each tumor composition. Statistical and sensitivity analyses were performed using standard packages in Matlab.

ACKNOWLEDGEMENTS
We thank Dane Wittrup and Leona Samson for their insightful comments. Funding was provided by Integrative Cancer Biology Program Grant U54-CA112967-06 (to M.T.H. and D.A.L.). B.Z. is supported by the NIH/NIGMS Interdepartmental Biotechnology Training Program 5T32GM008334.

AUTHOR CONTRIBUTIONS
B.Z. performed mathematical calculations and computational modeling. B. Z., M.T.H., and D.A.L. wrote the manuscript.
REFERENCES

(2014).

Figure 1. **Schematic of computational model.** *A.* Current clinical practice for diagnosis can result in sampling bias, with analysis based on only/mostly the predominant subpopulation. As such, a key question we try to address here is how intratumor heterogeneity (and specifically consideration of the entire heterogeneity versus just the predominant subpopulation) affects the resulting optimized drug combinations. *B.* To examine the effects of intratumor heterogeneity on optimal drug combinations, Monte Carlo sampling was applied to sample 10,000 heterogeneous tumor compositions. For each tumor composition, we mathematically optimized for a set of drug combinations. Statistical and sensitivity analyses were then applied on the resulting sampling results. *C.* Schematic for the optimization model. For a given tumor composition, drug combinations were optimized to maximize efficacy and minimize toxicity using a multi-objective optimization approach, scalarized using augmented weighted Tchebycheff method, and solved iteratively using linear programming (see Supplementary Text). Additional drug or tumor properties (derived from prior knowledge or machine learning) can also be incorporated into this framework. Solving the multi-objective optimization model leads to a set of solutions, or a Pareto optimal set, on the Pareto frontier, which represents a surface for which any increase in one objective results in a decrease in the other objective.
Figure 2. **Intratumor heterogeneity linearizes Pareto frontier and homogenizes drug combination efficacy.**

A. Representative objective space showing the Pareto frontier and the tradeoff between toxicity and efficacy for a homogeneous population of murine $E_{\mu}$-myc; $p19^{kip1}$ lymphoma cells infected with shp53 hairpin. Plot generated based on optimization described in Fig. 1C, using efficacy data comprised of efficacy of single drug for individual subpopulations (Fig. S1) and a symmetric toxicity profile for each drug. B. Representative objective space showing effects of heterogeneous population containing a predominant shp53 subpopulation and 13 other minor subpopulations on the Pareto frontier. C-D. Representative objective and solution space (shown for the same population as in (B)), with a maximum toxicity constraint of 6, which is set based on the number of drugs present in commonly used combination regimens. In the context of the mathematical model, this refers to the value for the parameter $a_{\text{max}}$ (see Supplementary Text). Compromise solution (colored green) refers to the point closest to the utopia based on $L_1$ norm distance metric.
Figure 3. Optimal drugs dominate at higher tumor complexity. A. Frequency distribution of component drugs in optimal drug combinations based on Monte Carlo sampling results, specifically for 6-drug combinations (regimen 6) and optimized using efficacy data (Fig. S1) and symmetric toxicity profile. Analyses shown in (B)-(C) are also based on the 6-drug combination. B. Frequency of component drugs as a function of tumor complexity (i.e. number of subpopulations in the heterogeneous tumor). The trend for each scatter plot is quantified with spearman correlation, the values of which are shown below the corresponding component drug title. C. Heatmap of the spearman correlations (in (B)) between component drug frequency and tumor complexity. Together (B) and (C) reveals that select set of drugs (e.g. rapamycin, 5-FU, vincristine, sunitinib) are strongly dependent on tumor complexity. D. Spearman correlation of component drug frequencies across different drug regimens in the Pareto optimal set. The high correlation illustrates that the frequency distribution of drugs for the 6-drug combination (shown in (A)) is comparable to other regimens (with N-drug combinations).
Figure 4. Drug optimization based on consideration of only the predominant subpopulation abrogates drug dominance at higher tumor complexity. Statistical analyses are formatted similar to Fig 3. However, here the analyses are based on drug optimization by consideration of only the predominant subpopulation in each heterogeneous tumor population. A. Frequency distribution of component drugs in optimal drug combinations, based on Monte Carlo sampling results, specifically for 6-drug combinations. Analyses shown in (B)-(C) are also based on the 6-drug combination. B. Frequency of component drugs as a function of tumor complexity. The trend for each scatter plot is quantified with spearman correlation, the values of which are shown below the corresponding component drug title. C. Heatmap of the spearman correlations (in (B)) between component drug frequency and tumor complexity. Together (B) and (C) reveals that in contrast to Fig. 3B and 3C, the drugs no longer depend on tumor complexity when only the predominant subpopulation is considered in the drug optimization. D. Frequency of component drugs in relation to other drug regimens in the Pareto optimal set. In contrast to Fig. 3C, the lower correlation values suggest that the drug frequencies are less consistent across the different regimens.
Different drug combination optimized based on entire heterogeneity
Same drug combination versus predominant subpop only

A. Breakdown of the optimal drug combination regimens for 10,000 Monte Carlo sampled heterogeneous tumor populations – showing proportions of solutions that are the same and different depending on whether the entire heterogeneity is considered versus just the predominant subpopulation.

B. Distribution of the difference in efficacy between drug combinations optimized based on entire heterogeneity versus just predominant subpopulation, for the 6-drug combination.

C. For each regimen in the Pareto optimal set, breakdown of solutions that are different based on the two optimization approaches to show the proportions for which drug combinations optimized based on entire heterogeneity still contains the single-best drug for the predominant subpopulation.

Figure 5. Drug combinations optimized based on consideration of entire heterogeneity can be non-intuitive.
Figure 6. Sensitivity analysis reveals optimal drugs are most robust and on average most efficacious. A. Point-biserial correlation ($r_{pb}$) to show the dependence of component drug choice on subpopulation. The drugs are ordered according to their frequency in optimal drug combination (Fig. 3A). B. Distribution of point-biserial correlations for each component drug. C. Correlation matrix of various drug characteristics. Component drug frequency is strongly positively correlated with mean efficacy and negatively correlated with kurtosis of $r_{pb}$, a metric used here for robustness.
Figure S1. Heatmap of efficacy data on effects of diverse drugs on subpopulations. Efficacy dataset\textsuperscript{26,29} comprised of 21 drugs and 30 subpopulations. The data was acquired based on an \textit{in vitro} competition assay, for which murine $E_{\mu}$-\textit{myc}; \textit{p19}$^{arf/-}$ lymphoma cells were partially infected with a shRNA targeting a gene of interest (and also becomes GFP+), treated with the indicated drug, and analyzed on a flow cytometry at 72 h for the relative GFP+ subpopulations. Enrichment/depletion of the GFP+ subpopulation relative to the parental were quantified in a log transformed resistance index (log2RI, see Supplemental Text), with positive and negative values representing enrichment and depletion, respectively. shCtrl refers to an empty vector control. The values presented in this figure are log2RI scaled to [0, $\infty$). Hierarchical clustering was performed with average linkage and Euclidean pairwise distance metric.
Figure S2. Pareto frontier and other terminologies in multi-objective optimization. A hypothetical example of the objective space showing the tradeoff between efficacy and toxicity, which we aim to maximize and minimize, respectively. Unlike a single objective optimization for which there exists a single global maximum or minimum, the solutions to multi-objective optimization consist of a Pareto optimal set, comprised of a set of solutions for which an increase in one objective results in a decrease in one or more objectives. This represents the inherent tradeoffs among objectives. The solutions in Pareto optimal set are non-dominating, meaning that there exist no other points such that all objectives are the same or superior, with at least one objective that is superior – otherwise the point is dominated. A weakly Pareto optimal set is a relaxed definition and also contains points for which objectives can be the same, as long as at least one objective is superior. The set of points in the Pareto optimal set constitute the Pareto frontier. Nonconvex region at the frontier, if exists, can be difficult to solve using weighted global criterion methods (and cannot be solved using the simple weighted sum approach). This motivated the use of the augmented weighted Tchebycheff method for scalarization (see Supplementary Text).
Figure S3. $p$ value selection for use in augmented weighted Tchebycheff method. Multi-objective optimization was scalarized using the augmented weighted Tchebycheff method (see Supplementary Text). A specific value of $p$ has to be used to ensure the necessary and sufficient condition for Pareto optimality. Shown in figure are the objective spaces for different $p$ values optimized on a homogeneous population of shATM with symmetric toxicity. A $p$ value of 10e-6 was chosen for subsequent optimization procedures.
Figure S4. Heterogeneity linearizes Pareto frontier. Objective space for (A) a homogeneous population comprised of 100% shChk2 and (B) a heterogeneous population comprised of a predominant subpopulation of 55% shChk2 and 25 other minor subpopulations. Multi-objective optimization was performed using the efficacy data (Fig. S1) and a symmetric toxicity profile. Compromise solution (colored green) refers to the point closest to the utopia based on $L_1$ norm distance metric.
Figure S5. Minimal correlation among drugs over populations. Matthews correlation of component
drugs, using drug choice (a binary value) for the 10,000 heterogeneous tumor populations (n by 10,000 matrix, n = number of drugs; any drugs with zero frequency were excluded because correlation would be undefined) with regimens optimized based on consideration of (A) the entire heterogeneity or (B) just the predominant subpopulation. Matthews correlation was calculated because correlation was between two binary variables.
Figure 5c. Tumor complexity is a crucial factor in drug frequency based on partial correlation.
**analysis.** (A) Spearman correlation of drugs over tumor complexity for the different regimens. Correlation was calculated for the 10,000 populations, grouped by number of subpopulations in each population (n by 30 matrix, n = number of drugs; any drugs with zero frequency were excluded because correlation would be undefined). (B) Partial spearman correlation of drugs, showing that much of the correlation is abrogated after controlling for tumor complexity.
Figure S7. Tumor complexity is no longer a critical factor if drug optimization only considers the
predominant subpopulation. (A) Spearman correlation of drugs over tumor complexity for the different regimens from the 10,000 Monte Carlo sampled tumor populations, with drugs optimized based on only the predominant subpopulation. (B) Partial spearman correlation of drugs after controlling for tumor complexity. In contrast to Fig. S7, correlation is comparable before and after taking tumor complexity into account.
Figure S8. A set of component drugs dominates at increasing tumor complexity when entire
heterogeneity is considered in optimization. Plots of frequency of component drug versus tumor complexity (i.e. number of subpopulations in heterogeneous tumor) for different regimens along Pareto frontier, generated from Monte Carlo sampling of 10,000 heterogeneous tumor populations, with drug combinations optimized based on efficacy data (Fig. S1) and a symmetric toxicity profile for the drugs. Optimization was performed based on consideration of all subpopulations in each heterogeneous tumor. Number below drug title shows the spearman correlation between drug frequency and tumor complexity.
Figure S9. Component drug dominance is when only the predominant subpopulation is considered.
**in optimization.** Plots of frequency of component drug versus tumor complexity (i.e. number of subpopulations in heterogeneous tumor) for different regimens along Pareto frontier, generated from Monte Carlo sampling of 10,000 heterogeneous tumor populations, with drug combinations optimized based on efficacy data (Fig. S1) and a symmetric toxicity profile for the drugs. Optimization was performed based on consideration of only the predominant subpopulation in each heterogeneous tumor. Number below drug title shows the spearman correlation between drug frequency and tumor complexity.
Figure S10. Drug combinations based on consideration of entire heterogeneity can be non-intuitive. (A) Breakdown of the optimized drug combinations for the 10,000 Monte Carlo sampled heterogeneous tumor populations – showing proportions of solutions with the same or different drug combination depending on whether the entire heterogeneity versus just the predominant or second largest subpopulation is considered. Similar analysis as that in Fig. 5, but now also examining drug combinations optimized based on the second largest subpopulation. (B) Additional breakdown of the drug combinations that are different based on the two optimization approaches – showing proportion of solutions still containing the single-best drug for the predominant and second largest subpopulations.
Figure S11. Sampling distribution of the 10,000 Monte Carlo sampled populations. Breakdown of the 10,000 Monte Carlo sampled populations into the predominant and second largest subpopulations. The heatmap results in a triangular profile because by definition the second largest subpopulation will always be less than or equal to the predominant subpopulation proportions. Some of the boundary cases (i.e. when the predominant and second largest subpopulation proportions are close) are sparsely sampled because the sampling was drawn from a uniform distribution first on the number of subpopulations in each tumor population. Oversampling of the boundary cases would bias toward smaller tumor sizes. This was overcome with additional sampling specifically near the boundaries (see Fig. S13).
Figure S12. Conclusions on non-intuitive drug combinations are robust when examining boundary cases. Breakdown of the optimized drug combinations for the 10,000 Monte Carlo sampled tumor populations – with 1) populations decomposed by the predominant subpopulation proportion versus the
difference in proportion between predominant and second largest subpopulation and 2) showing (A) fractions of drug solutions that are different based on the two optimization approaches (i.e. based on the entire heterogeneity versus just the predominant subpopulation). (B) Additional breakdown of (A) of the drug combinations that differ, the fraction of solutions that are non-intuitive (i.e. does not contain the single best drug for the predominant subpopulation). (C) and (D) Similar analyses as in (A) and (B), respectively, but with comparison between drug combinations optimized based on entire heterogeneity and just the predominant or second largest subpopulation.
A different drug combination optimized based on entire heterogeneity versus predominant subpopulation only.

Same drug combination.

Of the drug combinations that differ:
- Contains single-best drug for predominant subpopulation.
- Does not contain single-best drug for predominant subpopulation.

Frequency

Similarities in % between predominant and second largest subpopulations.

Cell-size: A, B, C, D, E.

 remaining on target subpopulations.
Different drug combination optimized based on entire heterogeneity versus predominant subpopulation only

Of the drug combinations that differ:
- Contains single-best drug for predominant or second largest subpopulation
- Does not contain single-best drug for predominant or second largest subpopulation

Figure S13. Additional Monte Carlo sampling at boundaries further revealed robustness of conclusions. The boundaries (i.e. small difference in proportions between the predominant largest and second largest subpopulations) were examined with additional Monte Carlo sampling, for a total of 13,050 heterogeneous tumor populations. Specifically, the first two subpopulation proportions were fixed with all possible combinations of 40%, 45%, and 50% predominant subpopulation and 1%, 3%, 5%, 7%, 9% difference between predominant subpopulation and second largest subpopulation (to be termed ‘deviance’). The choice for the first two subpopulations was enumerated across all \( \binom{30}{2} \) possible combinations. For each population, the remaining subpopulations were randomly drawn such that the summation of all the subpopulations is 100%. Subsequently, for each predominant subpopulation proportion and deviance, we examined the comparison between drug combinations optimized based on (A) the entire heterogeneity versus just the predominant subpopulation, and (B) the entire heterogeneity versus just the predominant or second largest subpopulations. The dotted gray line in the lower left corner plot of predominant subpopulation proportion versus deviance represents the theoretical limit for the
upper and lower bound (because the population must sum to 100% and the second largest subpopulation cannot be lower than 0%).
Figure S14. Representative objective and solution space under optimization with actual (asymmetric) toxicity. Plots of objective (A) and solution (B) space for optimization performed on the same 10,000 heterogeneous tumors previously sampled using the efficacy (Fig. S1) and toxicity (Fig. 7A) data, with maximum toxicity constraint of 6 for each side effect. Pink circles in the objective space refer to groupings of solutions sets with the same number of drugs in the combination.
Figure S15. Toxicity constraints can be incorporated in the multi-objective optimization. A. Curated binary toxicity matrix, A (see Supplementary Text for mathematical model), containing dose limiting toxicities for each component drug. B-C. Distribution of number of overlaps in toxicity for myelosuppression and gastrointestinal effects. A Gaussian distribution was fit to the underlying data (Fig. S16). An optimization with consideration of actual toxicity profile for each drug decreases overlapping toxicity in optimal drug combination, as compared to using symmetric toxicity profile. Toxicity is an additional consideration that can be thus incorporated into the multi-objective optimization framework.
Myelosuppression

Asymmetric toxicity

Symmetric toxicity

Gastrointestinal effects

Asymmetric toxicity

Symmetric toxicity

Figure S16. Distribution of toxicity overlaps under optimization of symmetric and actual (asymmetric) toxicity. Histograms of the toxicity overlaps, used for Gaussian distribution fits to generate Fig. S15B and S15C. Asymmetric toxicity refers to the use of actual toxicity profile (Fig. S15A). Optimizations based on these two toxicity profiles were performed using the same 10,000 Monte Carlo sampled heterogeneous tumor populations, used throughout this study.
Chapter IV

Exploiting temporal collateral sensitivity in tumor clonal evolution


Supplemental Tables will be available online.
**ABSTRACT**

The prevailing approach to addressing secondary drug resistance in cancer focuses on treating the resistance mechanisms at relapse. However, the dynamic nature of clonal evolution, along with potential fitness costs and cost compensations, may present exploitable vulnerabilities; a notion that we term 'temporal collateral sensitivity'. Using a combined pharmacological screen and drug resistance selection approach in a murine model of Ph+ acute lymphoblastic leukemia, we indeed find that temporal and/or persistent collateral sensitivity to non-classical BCR-ABL1 drugs arises in emergent tumor subpopulations during the evolution of resistance toward initial treatment with BCR-ABL1 targeted inhibitors. We determined the sensitization mechanism via genotypic, phenotypic, signaling, and binding measurements in combination with computational models, and demonstrated significant overall survival extension in mice. Additional stochastic mathematical models and small molecule screens extended our insights, indicating the value of focusing on evolutionary trajectories and pharmacological profiles to identify new strategies to treat dynamic tumor vulnerabilities.
INTRODUCTION

Collateral sensitivity describes a type of synthetic lethality that has been explored in cancer and infectious diseases for over forty years. Intrinsic to this concept is an evolutionary trade-off – where resistance toward a drug or drugs comes at the expense of sensitivity to other drugs. This phenomenon has spurred efforts to screen chemoresistant cell lines against a panel of drugs for collateral sensitivity and resistance (1, 2). Additionally, several recent high-throughput evolutionary experiments have attempted to build a collateral sensitivity networks using *E. coli* treated with 10-20 antibiotics, with the goal of designing drug cycling regimens that select against drug resistance (3–5). Evolutionary trade-offs have also been investigated for drug combinations (6–8), and have been utilized for potential control of subsequent tumor cell evolutionary trajectories (9–11).

In the field of cancer treatment, drug resistance studies have traditionally been focused on mechanism of resistance at the end of drug selection experiments. However, our understanding of intratumoral heterogeneity and clonal selection is increasingly revealing that tumor evolution is a dynamic process. Recent sequencing efforts have revealed extensive branched clonal evolution during tumor progression (12, 13), and matched samples prior and post treatment often enrich a pre-existing subclone toward dominance at relapse (14, 15). Such studies have also been recapitulated in *in vitro* settings with pre-existing resistant subclones estimated in one study at 0.001-0.05% of the parental population (16).

As with the bacterial antibiotics system, these evolutionary processes can sometimes present evolutionary trade-offs. Fitness costs of resistance have been extensively studied in bacteria, with findings that reduced fitness can in some cases concomitantly lead to acquisition of subsequent mutation(s) for cost compensation (17, 18). Therefore, we posited that there could be intermediate states during tumor clonal evolutionary progression that present persistent or temporal vulnerabilities. A conceptual illustration of our hypothesis is shown in Figure 1, where tumor genotype characteristics are represented on two abscissa axes of variation, such as could be the case for two independent gene mutations. “Fitness” on the ordinate axis is essentially the reciprocal of efficacy under whichever environmental conditions – such as a drug treatment – the tumor is dynamically evolving in. As a tumor becomes increasingly populated by cells resistant to treatment with an initial drug A (up-hill mountain in panel A), there can be drugs from distinct drug categories that are inactive (Fig 1B) or collaterally sensitizing to the terminal resistant stage of clonal evolution (Fig 1C). However, it is also conceivable that certain genotype variations (e.g. on-target mutations and/or alterations in signaling pathways) would render at least a proportion of the evolving tumor more susceptible to a different drug D (down-hill valley in panel D). This situation in theory could lead to a treatment regimen with drug D following drug A during a restricted time-window.
producing overall increased treatment efficacy – a notion that can be termed ‘temporal collateral sensitivity’.

In this study, we report an experimental validation of temporal collateral sensitivity in a murine preclinical model of Philadelphia chromosome (Ph)-positive acute lymphoblastic leukemia (ALL). Current BCR-ABL1 (the fusion protein generated by the Philadelphia chromosome) inhibitors have dramatically improved the prognosis for chronic myeloid leukemia and to some degree Ph+ ALL. However, a major resistance mechanism for relapsed patients is the on-target BCR-ABL1 dependent mutations in the kinase domain, including the most common gatekeeper T3151 mutation that confers resistance to four of the five FDA-approved drugs (19–21). Even then, compound mutations (i.e. more than one mutation in the same BCR-ABL1 allele) remains a challenge (22, 23). Through a joint drug resistance selection and pharmacological profile, here we discover a particular strong and robust collateral sensitivity to non-classical BCR-ABL1 inhibitors (i.e. not known and used to target BCR-ABL1) crizotinib, foretinib, cabozantinib, and vandetanib for Ph+ ALL cells at intermediate stages of clonal evolution. We show that this effect is driven by the occurrence of the BCR-ABL1 V299L single mutation, before its continued evolution toward V299L compound mutants. Molecular studies suggest an on-target ABL1 inhibition exclusive to V299L (and a subset of V299L compound mutants). Most importantly, the mechanism is distinct from drug synergy as a result of temporal cellular rewiring, but rather via selection of stable collaterally sensitive clones. Thus, our work suggests new strategies for rationalizing drug (or drug combination) design beyond the focus on traditional terminal stages of clonal evolution and emphasizes an understanding of the underlying evolutionary trajectories and intermediate stages to fully exploit tumor vulnerabilities.

RESULTS

Drugs with fitness landscapes can affect clonal intermediates and diversify subpopulation trajectories

To explore the conceptual idea of temporal collateral drug sensitivity arising from tumor subpopulation dynamic evolution, we first developed a minimal stochastic branching process based mathematical model to examine the effects of distinct predefined fitness landscapes on clonal evolution (see Methods and Supplemental Methods for details).

Under the premise of a step-wise clonal evolution, we assumed a fitness landscape built based on multivariate Gaussian distributions with an intermediate and terminal stage (Fig 1A). In the presence of drug, the simulation revealed the treatment-naïve initial population would expectably evolve toward higher fitness (Supp Fig 1A and 2A). We next hypothesized the existence of an evolutionary trade-off at
the intermediate stage of clonal evolution – specifically exhibiting a suboptimal fitness to another drug D (Fig 1D). As such, during a clonal evolution under drug A selection, there exists a treatment window (‘temporary collateral sensitivity’) during which a switch to drug D will lead to a lower fitness of the resulting population (Supp Fig 1B).

Using this same model, we asked what the consequences would be if we first use drug A for resistance selection, followed by drug D – to exploit this temporal collateral sensitivity. Simulation results suggest that sequential drug switch and selection can change the propensities of clonal evolution, and lead to divergent trajectories (Supp Fig 1C and 2B). Specifically, continued selection in drug D led to distinct trajectories that although all have high fitness for drug D (Supp Fig 1D), now have populations with diverse fitness to drug A (Supp Fig 1C). Thus, examining drugs with distinct fitness landscapes can potentially expose a treatment window with temporal vulnerabilities during clonal evolution, and the choice of sequential drug selection can change/diversify subsequent clonal trajectories. Of note, the mathematical model was based on our preconception of a step-wise fitness landscape. The model can be modified and/or extended with incorporation of random field models of fitness landscapes (e.g. NK model) and/or examination of different drug schedules. However, here we chose next to turn to an experimental system to screen and validate this conceptual idea more concretely.

A pharmacological screen identifies intermediate BCR-ABL1 V299L populations with collateral sensitivity to crizotinib, foretinib, cabozantinib, and vandetanib

Motivated by the results from the stochastic branching process model, we examined combined pharmacological profiles and in vitro drug resistance selection in a murine preclinical model of Ph+ acute lymphoblastic leukemia (24, 25). Ph+ ALL is a particularly relevant and tractable model – in contrast to chronic myeloid leukemia (CML), Ph+ ALL is an aggressive disease and has rapid relapse following treatment with frequent selection of BCR-ABL1 kinase domain mutations (20, 26, 27). Here, we derived resistant cell lines with dose escalating concentrations of BCR-ABL1 inhibitors (Fig 2A). Each derived cell line (totaling close to 180 cell lines) was screened across a diverse set of small molecule inhibitors. In our preliminary screen, through several independent dasatinib selections, we observed initial resistance to dasatinib and cross-resistance to bosutinib (Fig 2B). Strikingly, we also observed strong sensitization (i.e. collateral sensitivity) to crizotinib, foretinib, vandetanib, and cabozantinib. These drugs are typically recognized as cMET and/or VEGFR inhibitors, but the sensitization we found did not appear to be generalizable across all such targeted drugs examined – suggesting a mechanism of action beyond these known canonical targets.
We continued with multiple rounds of independent selection to first confirm this was indeed a robust phenotype. With over ten independent selection series, we observed a consistent sensitization to these four small molecules upon the development of cross-resistance to dasatinib and bosutinib. Most interestingly, as the cells continued to evolve toward cross-resistance to imatinib, nilotinib, and at times ponatinib, the magnitude of collateral sensitivity diminished (Fig 2C). This was particularly prominent for crizotinib, foretinib, and cabozantinib. Vandetanib appeared to largely retain its collateral sensitivity over the entirety of clonal evolution. In addition to this sensitization, we also observed other distinct patterns of collateral resistance and sensitivity, even within the same drug through independent selections. This highlights the stochastic nature of this process. As would be the case from Luria and Delbruck’s fluctuation analyses, this variability suggests selection of distinct clones for outgrowth, which would lead to subsequent different phenotypic pharmacological response patterns.

Since on-target ABL1 kinase domain mutations are a common mechanism of resistance to BCR-ABL1 inhibitors, we PCR-amplified and Sanger sequenced the kinase domain of ABL1. We observed a perfect concordance between the collateral sensitivity to crizotinib, foretinib, cabozantinib, and vandetanib, and the presence of a single V299L mutation in ABL1 (Fig 2C and 2D). Remarkably, the dasatinib/bosutinib-resistant cells containing the V299L mutation further evolved to develop V299L compound mutations under continued drug selection. The emergence of V299L compound mutations was coincident with the reduced collateral sensitization (Fig 2C). Although variant calls from RNA-seq of select cell lines containing BCR-ABL1 V299L or V299L compound mutants revealed additional passenger SNVs and indels (Supp Fig 3), the only mutation shared among cell lines with the sensitization phenotype and that went from 0% to 100% variant allele frequency was c.895G>C, leading to BCR-ABL1 V299L (Supp Fig 4 and Supp Table S1).

To further confirm that BCR-ABL1 V299L is the causative mutation in sensitizing cells to these four small molecules, we performed dose response experiments using isogenic Ba/F3 cell lines. While BCR-ABL1 V299L expectedly conferred resistance to dasatinib and bosutinib, we observed strong and robust sensitization to crizotinib, foretinib, cabozantinib, and vandetanib (Fig 2E).

**BCR-ABL1 V299L is pre-existing and selected for during drug treatment**

We next investigated whether the BCR-ABL1 V299L and/or V299L compound mutants were pre-existing in the parental population. Since both Sanger sequencing and variant-calls from RNA-seq did not detect BCR-ABL1 V299L in the parental population, the actual estimate would at most exist at a level that is below the detection limit of these technologies (~10% for Sanger sequencing, and 0.1-1% for NGS
technologies (28)). Here we developed a stochastic mathematical model based on nonhomogeneous continuous-time multi-type birth-death process (see Methods and Supplemental Methods). The model incorporates the background mutation rate, birth and death rates of the individual subpopulations (derived based on experimentally determined net growth rates), and death rate in the presence of drug (derived from dose response curves). We performed Monte Carlo simulations with various combinations of parameter values, taking into account the uncertainty and initially zero or some non-zero value of BCR-ABL1 V299L or V299L compound mutant subpopulation at the beginning of the simulation. The only set of parameters that best explain our experimentally observed kinetics (9 days for initial outgrowth, and V299L in going from < 0.1% to >99.9% at day 9) is the pre-existence of a BCR-ABL1 V299L subpopulation at 0.0082% in the parental population (Fig 3A-C and Supp Fig. S5). Furthermore, the model predicts that the BCR-ABL1 V299L compound mutant could exist either at zero or a very minor (i.e. < 0.0006%) percentage at the beginning of the selection experiments (Fig 3D and Supp Fig. S5). In addition, our sensitivity analysis demonstrates that this result is robust and largely independent of background mutation rates and birth/death rates (Fig 3E). The major determinant of the kinetics and fractional appearances of subclones is the initial subpopulation size.

Using the estimated pre-existing fractions of BCR-ABL1 V299L and V299L compound subpopulations, we next determined the approximate treatment window for exploiting collateral sensitivity, which was driven by the predominance of V299L single mutation at intermediate stages of clonal evolution. We performed mathematical modeling based on a system of ordinary differential equations to fully simulate our selection experiments – with automatic dose escalation upon outgrowth (Fig 3F and 3G). The model revealed an approximate range of 1 to 3 weeks where BCR-ABL1 V299L was the dominating subpopulation fraction and, as such, conferring a persistent sensitization to drugs such as foretinib (Fig. 3H and 3I).

Another aspect with regards to timing and clinical management is the scheduling of drug combinations. To examine this question, we simulated tumor kinetics under different drug schedules (single, concurrent, or alternating of dasatinib and foretinib). We observed that when the pre-existing V299L subpopulation is small, the most synergistic combination was that of an alternating regimen, especially if the concurrent treatment requires dose reduction (Supp Fig S6A). Here foretinib would be effective only after the enrichment of the V299L subpopulation following dasatinib treatment. However, once V299L reached a substantial fraction, the therapeutic benefit of an alternating treatment is limited as we risk the outgrowth of the BCR-ABL1 WT subpopulation during the foretinib cycle, or the outgrowth of V299L compound from V299L subpopulations during the dasatinib cycle (Supp Fig S6B). As such, the efficacious strategy
remains to be a concurrent treatment of dasatinib and foretinib, or alternating treatments of a duration that can control the outgrowth of the BCR-ABL1 WT and V299L effectively.

**BCR-ABL1 V299L confers collateral sensitivity through an on-target BCR-ABL1 inhibition**

Our demonstrated efficacy in isogenic Ba/F3 cell lines suggest that the collateral sensitivity is likely to act through an on-target BCR-ABL1 inhibition. To confirm, we made various phenotypic and signaling measurements. BCR-ABL1 promotes cell cycle entry (G1-to-S phase transition) and its inhibition results in G1 cell cycle arrest (29–31). As such we posited that a similar cell cycle profile would provide, albeit a downstream read-out, evidence of potential on-target ABL1 inhibition. Therefore, we examined cell cycle profiles of BCR-ABL1 WT and BCR-ABL1 V299L cells upon drug treatment. As expected, we observed an induction of G1 arrest in both WT and V299L cell lines upon treatment with BCR-ABL1 inhibitors (Fig 4A and Supp Fig S7). In addition, as was expected from previous dose response curves, higher concentrations of dasatinib and bosutinib were required for G1 arrest in the presence of V299L. However, lower concentrations of vandetanib, cabozantinib, crizotinib, and foretinib was sufficient to induce a G1 arrest in BCR-ABL1 V299L cells (Fig 4A). Furthermore, crizotinib and foretinib appeared to exhibit off-targets at high concentrations, leading to a prominent G2/M arrest in both BCR-ABL1 WT and V299L cell lines. This was consistent with a previous report of G2/M arrest (and mitotic catastrophe) induced by foretinib in CML K562 cells (32), albeit the exact mechanism mediating this effect remains unknown.

Since Stat5 is a substrate of BCR-ABL1 and a surrogate biomarker of BCR-ABL1 activity (21), we also examined phospho-Stat5 levels and cleaved PARP, as a marker for apoptosis via flow cytometry. As expected, we observed down-regulation of pStat5 upon treatment with BCR-ABL1 inhibitors, and again with a higher required concentration for dasatinib/bosutinib in V299L cell lines (Fig 4B). Although we observed no down-regulation of pStat5 in BCR-ABL1 WT cell lines upon treatment with vandetanib, cabozantinib, crizotinib, and foretinib, the occurrence of down-regulation in pStat5 in V299L cell lines suggests an on-target ABL1 inhibition in the presence of BCR-ABL1 V299L. An increased cPARP level without a strong pStat5 inhibition at high concentrations of crizotinib and foretinib in both BCR-ABL1 WT and V299L cell lines is consistent with the earlier cell cycle profiles suggesting an off-target mechanism of action at the higher doses.

To show direct inhibition of ABL1, we further performed *in vitro* kinase assays with purified recombinant active ABL1 WT and ABL1 V299L mutant. We observed preferential inhibition of kinase activity for V299L mutant compared to WT for the four small molecules – crizotinib, foretinib, cabozantinib, and vandetanib (Fig 4C and Supp Fig S8). Given the particularly large fold change in preferential inhibition
for vandetanib against ABL1 V299L based on our kinase and viability assays, we also performed computational docking studies. Intriguingly, bosutinib and vandetanib share close chemical structures: quinoline (for bosutinib) or quinazoline (for vandetanib) group for occupying the adenine pocket, a substituted aniline group for occupying the hydrophobic pocket, and a long-chain extended into the solvent region for increased solubility (Supp Fig 9A). Our docking results support this insight, with vandetanib binding to ABL1 WT and V299L in a similar conformation to that for bosutinib (Fig 4D and 4E). BCR-ABL1 inhibitor bosutinib forms a hydrogen bond to the hinge region of ABL1 (a feature shared by most kinase inhibitors), and via its nitrile group van der Waals contacts with T315 and V299 and water-mediated hydrogen bond network to the DFG motif (33, 34). Mutation V299L results in steric hindrance to the nitrile group, and the likely cause of clinically observed resistance. Most notably, docking of vandetanib to V299L kinase domain suggests leucine capable of making an additional non-polar contact with vandetanib (Fig 4E and Supp Fig 9B). Energetic calculations suggest vandetanib in complex with V299L mutant is stabilized by approximately ten fold through improved packing – primarily via interaction between V299L and the quinazoline group of vandetanib.

To rule out any off-target effects, we examined RNA-seq differential expression of our parental, V299L-, and V299L compound-containing cell lines. Unsupervised hierarchical clustering showed the expected groupings with respect to the cell lines’ ABL1 mutational status (Supp Fig S10). Although there were limited differentially expressed genes among the cell lines (Supp Fig S11), gene set enrichment analyses did not reveal any statistically significant phenotypic or functional categories. This suggests that V299L (and V299L compounds) does not affect cellular state. Next, we assessed the possibility that this mutant would lead to substrate specificity changes downstream BCR-ABL1. We integrated results from a recently-developed algorithm named KINspect (35) that identifies those residues most and least likely to contribute to substrate specificity by exploring millions of different specificity models. The algorithm confirmed that V299L, with a KINspect specificity score of 0.05, is unlikely to disrupt or significantly change substrate specificity (Supp Table S2). Taken together, these data indicate that the collateral sensitization conferred by BCR-ABL1 V299L is due to on-target ABL1 inhibition.

Collaterally sensitive drugs against BCR-ABL1 V299L exhibit in vivo efficacy
To investigate whether the collateral sensitivity translates to in vivo efficacy, we tail-vein injected BCR-ABL1 WT or V299L cell lines into immunocompetent syngeneic recipient mice. Upon the cytological appearance of an initial tumor burden on day 10 post transplantation, we treated mice once daily with vehicle, imatinib, dasatinib, foretinib, or vandetanib. We first assessed the overall tumor burden using whole-mouse bioluminescence imaging pre- and post-treatment. As expected, we observed a reduction in
tumor burden upon treatment with imatinib and dasatinib in mice with WT BCR-ABL1 (Fig 5A and 5B). Foretinib and vandetanib had modest to no in vivo efficacy. However, in mice with BCR-ABL1 V299L, imatinib and dasatinib exhibited no antitumor responses, while treatment with foretinib or vandetanib led to a significant 3-log fold reduction in tumor burden compared to vehicle control (Fig 5B).

Animals transplanted with these Ph+ ALL cells also develop splenomegaly. Whereas treatment of WT BCR-ABL1 ALL-bearing mice with foretinib or vandetanib did not lead to any reduction in spleen size, mice bearing BCR-ABL1 V299L ALL exhibited a 3- to 4-fold reduction in spleen size upon treatment (Fig 5C and 5D). Of note, we observed a moderate fitness defect for BCR-ABL1 V299L cells relative to cells expressing WT BCR-ABL1. This was apparent both based on in vitro growth assays and in the disease onset for these tumors. As such, BCR-ABL1 WT mice were also sacrificed a few days early for imaging and assessment of spleen size. BCR-ABL1 mutants have been known to exhibit different fitness in the absence of drug, which may be due in part to moderate catalytic inefficiencies and/or downstream substrate specificity (36, 37).

We next evaluated the overall survival of treated animals. We performed once daily oral gavage of vehicle, imatinib, dasatinib, foretinib, or vandetanib for one week starting at day 10 post-transplantation with cells expressing either BCR-ABL1 WT or V299L. In mice with WT BCR-ABL1, we observed an increase in median survival of 4 and 5 days upon treatment with imatinib or dasatinib, respectively (Fig 5E, P < 0.001). Treatments with foretinib or vandetanib did not lead to a statistically significant increase in survival for mice bearing WT BCR-ABL1. In contrast, we observed a significant extension in median survival of 3 to 4 weeks when mice with BCR-ABL1 V299L were treated with foretinib or vandetanib. Most notably, a subset of the animals in these cohorts achieved long-term survival without relapse.

**Sequential drug combination selection reveals divergent clonal trajectories**

Given the relevance of collateral sensitivity to treatment response in this model, we next wondered how the choice of sequential drug selection affects the occurrence of temporal collateral sensitivity and clonal trajectories. To examine these questions, we performed a similar drug resistance selection experiment as before. However, as opposed to using a single drug at dose escalating concentrations, we treated cells with a single dose of drug A at an IC90 concentration, followed, after initial tumor cell regrowth, by a switch to drug B at dose escalating concentrations (Fig 6A).

Here, we again observed that, upon selection with dasatinib or bosutinib, the parental ALL cell line has a propensity toward the development of BCR-ABL1 V299L mutations (Fig 6B). A single round of
selection with the other non-ABL1 targeting small molecules led to zero or one mutations in the ABL1 kinase domain. In either case, the resulting pharmacological response revealed no pronounced phenotypes of collateral resistance or sensitivity.

Continued selection with BCR-ABL1 V299L ALL cells using dose-escalating concentrations of non-ABL1 targeting small molecules led to continued evolution toward V299L compound mutants. In some cases, the V299L compound (i.e. V299L/F317L) resulted in pronounced attenuation of collateral sensitivity, whereas in others (e.g. V299L/D276G or V299L/Q252H), the collateral sensitivity appeared to be retained (Fig 6C). Structural modeling and analysis shows that F317L independently weakens vandetanib binding through diminished packing interactions with the quinazoline group. In contrast, changes in Q252 and D276 are expected to have minimal effect – the former is solvent exposed and makes only weak intra-protein interactions and the latter is over 20 Å away from the active site. Together, these data suggest additional mutations in ABL1 can result in neutral or disruption of the stabilization previously created via V299L.

We next examined the effect of initial selection with the collaterally sensitive small molecules, followed by dose escalating selection with dasatinib. We observed that BCR-ABL1 V299L was no longer the dominating resistant population at outgrowth. Although all the mutations continue to be localized to the ABL1 kinase domain, we observed a more diverse set of trajectories, including BCR-ABL1 T315I (Fig 6C). Thus, the order in which cells are exposed to drugs/develop resistance can affect the propensities of cells toward specific resistance mechanisms. In particular, we still do observe V299L in these sequential selection experiments; however, this pre-selection of e.g. foretinib followed by dasatinib potentially diminished the pre-existing predominant V299L subpopulation and enabled the stochastic selection and outgrowth of other subclones.

Small molecule screen reveals compounds with diverse fitness landscapes across clonal evolution
We have thus far examined collateral sensitivity and resistance using select set of targeted and chemotherapeutics. We next wanted to investigate the diversity of fitness landscapes in a broader small molecule space. Capitalizing on the known trajectories of our parental ALL cell line in evolving toward V299L and subsequently V299L compound mutants upon dasatinib selection, we performed a small molecule screen of 391 compounds against the parental, BCR-ABL1 V299L, and V299L/E255K cell lines (corresponding to the initial, intermediate and terminal stage of clonal evolution) (Fig 7A).
We can schematically visualize the EC50s as they are mapped onto a fitness landscape, consisting of these three clonal stages. Of the positive controls, there are molecules such as bosutinib (Fig 7B) and imatinib (Fig 7C) where there is an escalating fitness with progressive stages of clonal evolution. In contrast, there are other drugs such as vandetanib and foretinib where intermediate and/or terminal stages consist of valleys instead of peaks in the fitness landscape – an illustration of collateral sensitivity. The other possibility is a valley for the intermediate and a peak for the terminal, as was observed for axitinib (Fig 7F) and crizotinib (Fig 7G). Interestingly, axitinib was recently reported to exhibit efficacy against BCR-ABL1 T315I (as well as V299L), through an on-target ABL1 inhibition (38). Studies here suggest that the V299L/E255K compound abrogates such efficacy for axitinib. These are among several small molecules that appeared to exhibit these distinct fitness landscapes, with effects ranging from persistent to temporal collateral resistance or sensitivity.

DISCUSSION

Our view of tumor clonal evolution as a dynamic process led us to propose the existence of additional vulnerabilities and evolutionary tradeoffs for exploitation – en route toward terminal drug resistance. Notably, using combined in vitro drug resistance selection and pharmacological profile experiments, we have uncovered a temporal collateral sensitivity phenomenon using a preclinical murine model of Ph+ ALL. Sequencing analysis in partnership with kinetic mathematical modeling revealed that this was driven by a selection of a pre-existing BCR-ABL1 V299L and subsequently BCR-ABL1 V299L compounds-containing stable clonal subpopulations. As such, this is quite distinct from temporal network rewiring mechanisms of drug synergy/sensitization. Our RNA-seq, phenotypic, signaling, binding, and docking analyses indicate that the sensitization to non-classical BCR-ABL1 inhibitors crizotinib, foretinib, vandetanib, and cabozantinib was driven by an on-target ABL1 inhibition in the presence of the V299L mutation. Clinically, V299L mutations have been observed in CML and Ph+ ALL patients, and occurs upon dasatinib or bosutinib treatment failure (at 5-7%) (39, 40). V299L compound mutations are also clinically observed (23), and in one study of CML patients, V299L was the second most common component of compound mutations (at 20%) (22).

The mechanism of action for this collateral sensitivity was surprising: a gained on-target ABL1 inhibition realized by new interaction forged through the initial ABL1 mutation. The selectivity toward a specific kinase mutant vis-à-vis WT is not unprecedented. Recently axitinib has been shown to be active against BCR-ABL1 T315I (and other BCR-ABL1 mutants, including V299L) (38). Here axitinib binds to the active site of ABL1 in a distinct conformation different from its known target VEGFR2. In contrast, pan-Aurora inhibitor danusertib binds to ABL1 T315I in similar conformation, with Ile315 mimicking Leu210
in Aurora A (41). Our findings and these studies suggest existing compounds may exhibit selectivity toward kinase mutants over WT that in some cases due to the mutation changing to an ‘on-target mimic’. This highlights opportunities for potential experimental and in silico screens of existing small molecules for mutant selectivity and possibly provide alternatives to existing drugs with known adverse side effects (e.g. ponatinib and nilotinib with vascular occlusive events (42)).

Capturing the time window for temporal collateral sensitivity can be challenging. However, for Ph+ ALL or CML, biopsies and quantitative analysis of BCR-ABL1 transcript levels are routinely utilized for monitoring disease progression. This provides the opportunity to strategize the timing for incorporation of non-classical BCR-ABL1 inhibitors for patients. In solid tumors and other hematopoietic cancers, utility of circulating tumor cells or DNA (albeit at limited resolution) may be used as additional modalities. Nevertheless, challenges remain for the identification and improved sensitivity of biomarker identification.

Ultimately, optimal clinical management will also require the incorporation and consideration of current standard of care. In the case for Ph+ ALL, current regimens include fractionated chemotherapy combinations, such as Hyper-CVAD with addition of a BCR-ABL1 inhibitor, and most often followed by stem cell transplantation. However, incorporation of tyrosine kinase inhibitors (TKIs) into chemotherapy regimens has not dramatically improved remission duration (27), suggesting a need for additional understanding on the effects of chemotherapy and TKIs and/or clinical management for this aggressive disease.

Our modeling presented here provides a conceptual understanding of temporal collateral sensitivity and estimations of pre-existing populations. Additional experimental and mathematical modeling will be useful to examine different drug scheduling strategies. For example, alternating treatment may be of value if cycling between classical BCR-ABL1 inhibitors and collaterally sensitive drugs may impede the continual evolution toward compound mutants. However, if a large wild-type fraction still persists after the initial log fold tumor reduction, a more practical strategy would involve concurrent treatments with inclusion of classical BCR-ABL1 inhibitors (or hyper-fractionation of the alternating regimen) for the effective management of disease. Mathematical modeling of these scenarios along with incorporation of various factors (e.g. pharmacokinetics, clonal interference/cooperation, microenvironments, etc.) and applications from control theory (43) and adaptive drug design (44) will inform scheduling implications in the context of temporal and/or persistent collateral sensitivity trade-offs.
Undoubtedly different parental cell lines and patients prior to diagnosis will exhibit different propensities toward resistance. This study reveals that there can potentially be a treatment window to exploit temporal collateral sensitivity at predictable clonal evolutionary trajectories. Furthermore, this work highlights an approach for combining extensive drug resistance selection experiments with pharmacological profiles to identify novel vulnerabilities during the course of tumor clonal evolution. In particular, this approach has broad applications for studies beyond tyrosine kinase inhibitors. For instance, multiple rounds of chemotherapy selection can result in distinct step-wise progression via multiple mutations. The pharmacological profiles would not only identify these functional clonal evolutionary states that would otherwise be difficult solely based on genome sequencing, but also the identification of specific vulnerabilities at such stages. It is worth noting that this mechanism does not have to be restricted to on-target mutations, but is also a potential consequence of dysregulation in compensatory pathways. Our Ph+ ALL model exhibits strong evolutionary constraints toward on-target point mutations as mechanism of resistance. Different cancer types and drug classes (including chemotherapy) will undoubtedly have distinct mechanisms. Nevertheless, similarity in evolutionary constraints – including prevalence of on-target kinase domain mutations in ALK (in non-small-cell lung cancer) (45) or cKIT (in gastrointestinal stromal tumor) (46) and known compensatory pathways (e.g. in EGFR, IGFR1, cMET, etc) presents opportunities to survey more densely their clonal evolutionary trajectories and potential intermittent vulnerabilities.

**EXPERIMENTAL AND COMPUTATIONAL PROCEDURES**

*Cell lines and chemicals*

Murine derived Ph+ALL cell line containing BCR-ABL1 p185 (24, 47) was cultured in RPMI medium supplemented with 10% FBS, 4 mM L-glutamine, 5 μM β-mercaptoethanol. The murine Ba/F3 parental cell line, received from the Druker lab, cultured in RPMI medium supplemented with 10% FBS and 10 ng/mL recombinant murine IL3 (R&D Systems). Mutant Ba/F3 expressing BCR-ABL1 wild-type or BCR-ABL1 V299L were generated as previously described (48). Mutant Ba/F3 were cultured in RPMI medium supplemented with 10% FBS. All cell lines were tested and shown to be free of *Mycoplasma* using PCR-based (ATCC) and biochemical-based (Lonza) methods. All drugs were obtained from LC Laboratories or Selleck Chemicals.

*Drug resistance selection*

Cells were plated at 0.5 million cells/mL and treated at desired drug concentrations. Cells were monitored each day. Upon outgrowth (i.e. cell density reaching 4-5 million cells/mL), cells were 1) frozen down, 2)
allowed to recover in no-drug medium, or 3) plated again at 0.5 million cells/mL and treated at the next drug dose. For cells in drug-selection condition, medium was changed every week. For cells recovering in no-drug medium, genomic DNA was extracted for PCR amplification and Sanger sequencing of the ABL1 kinase domain, and dose responses were performed. The stability of the cell line in no-drug medium was confirmed through both Sanger sequencing and dose responses, performed periodically up to 2 months post-recovery.

Statistical analyses
Statistical analyses were performed using Prism v5 (GraphPad) and R v3.2.0. Comparisons in tumor burden reduction were assessed using Mann-Whitney test. RNA-seq differential expression analyses were performed using DESeq package in R. Survival of mice was analyzed with Kaplan-Meier method with significance assessed using log-rank test.

Accession numbers
RNA-seq data have been deposited in Gene Expression Omnibus under the accession number GSE72910.

Further details on experimental procedures (proliferation assays, kinase assays, cell cycle and signaling measurements, PCR of ABL1 kinase domain, RNA-seq analyses, high-throughput small molecule screen, and in vivo studies) and computational methods (small molecule docking, KINspect calculation, and mathematical modeling) can be found in Supplemental Experimental and Computational Procedures.

AUTHOR CONTRIBUTIONS
B.Z., D.A.L., and M.T.H. designed research. B.Z. performed in vitro and in vivo experiments and computational modeling. J.C.S. made contributions to the ODE mathematical model. R.S. and B.T. performed small molecule docking. P.C. performed computational analyses to evaluate the likelihood of mutants to cause changes in kinase specificity using KINspect. J.R.P. provided reagents and conceptual contributions toward research design and data analysis. B.Z., D.A.L., and M.T.H. analyzed data and wrote the manuscript with input from all authors.

ACKNOWLEDGEMENTS
We thank Peter Bruno and Christian Braun for critical review and feedback on this manuscript, and Stuart Levine and Jie Wu from the MIT BioMicroCenter and Glen Paradis from the Koch Institute Flow Cytometry Core for services and advices. This work was supported by the Koch Institute Support (core)
Grant P30-CA14051 from the National Cancer Institute, the Integrative Cancer Biology Program Grant U54-CA112967 (to M.T.H., D.A.L., and B.T.), and the National Institutes of General Medical Sciences GM082209 (to B.T). B.Z. and R.S. are supported by the NIH/NIGMS Interdepartmental Biotechnology Training Program 5T32GM008334. P.C. is supported by a Ludwig Fund Postdoctoral Fellowship. B.Z. is a consultant for ARIAD Pharmaceuticals, Inc. J.R.P. is an employee and shareholder of ARIAD Pharmaceuticals, Inc. D.A.L. is a consultant/advisory board member of Merrimack Pharmaceuticals, Genentech, and Complete Genomics. Remaining authors declare no potential conflict of interests.
REFERENCES
19. S. Branford et al., Detection of BCR-ABL mutations in patients with CML treated with imatinib is virtually always accompanied by clinical resistance, and mutations in the ATP phosphate-binding loop (P-loop) are associated with a poor prognosis, Blood 102, 276–283 (2003).
22. J. S. Khorashad et al., BCR-ABL1 compound mutations in tyrosine kinase inhibitor-resistant


30. R. A. Gatenby, A. S. Silva, R. J. Gillies, B. R. Frieden, Adaptive therapy., Cancer Res. 69, 4894–
Figure 1. **Conceptual fitness landscapes with clonal intermediates.** Predefined fitness landscapes can be visualized with a z-axis corresponding to a fitness of the population under a given environmental condition, and x- and y- corresponding to a two-dimensional coordinate of the genotype of each subpopulation. The actual genotype can be in a high-dimensional space, but is explicitly represented here in two-dimensions. The fitness landscape for drug A is composed of two Gaussian peaks for intermediate and terminal stage. In contrast, at the location of the intermediate peak, the corresponding fitness landscape for drug B contains a valley with fitness. Initial population is a homogeneous population starting at a low fitness, as indicated by the white asterisk.
Figure 2. A pharmacological screen of each distinct evolutionary stage to identify persistent and temporal collateral resistance and sensitivity. (A). Schematic of experimental setup for drug resistance selection experiment (see Methods for details). Briefly, a murine derived Ph+ acute lymphoblastic leukemia cell line is treated at IC90 1x drug concentration. Upon recovery and outgrowth, the population is dose escalated to 2x the previous drug concentration. The derived cell line, mimicking a specific stage of the clonal evolution, was allowed to recover and profiled based on viability assays across a panel of targeted and chemotherapeutics. Each selection experiment terminates upon either no outgrowth at the
given drug concentration or until IC90 16x. (B) Preliminary drug selection experiment with DMSO control and three independent dasatinib selection at IC90 1x concentration, illustrating collateral sensitivity and resistance. (C) A complete overview of the drug selection experiments for vehicle, dasatinib, and bosutinib, showing diverse collateral resistance and sensitivity patterns. The black triangles illustrate each independent series of dose escalating concentrations, and hence also an indicator of time. The heatmap shows the log2 transform of the ratio of the EC50s for each drug of given cell line relative to parental cell line. Cell lines with DMSO control grown in parallel had similar EC50s as the parental. The kinase domain of ABL1 was also Sanger sequenced, the subpanel to the right of heatmap illustrates complete concurrence between the sensitization to crizotinib, foretinib, cabozantinib, and vandetanib and single mutational V299L in ABL1. (D) A representative Sanger sequencing of V299L. (E) Dose responses of BCR-ABL1 inhibitors and collaterally sensitive inhibitors crizotinib, foretinib, cabozantinib, and vandetanib in Ba/F3 isogenic parental, BCR-ABL1 WT, and BCR-ABL1 V299L cell lines. The sensitization was consistently observed and suggests that V299L is a causative determinant for the sensitization phenotype.
Figure 3. Mathematical models of tumor kinetics predicts pre-existing subpopulations and treatment window. (A-B) Representative stochastic birth/death model simulation results with Monte Carlo sampling of parameters for the first round of selection with dasatinib at IC90 1x concentration. The total population size is shown in (A) and the corresponding subpopulation fractions shown in (B). Simulation results were constrained to those fitting experimental observations (in terms of total population size and tumor composition) at day 0 and 9. Stochastic model predictions at ten evenly distributed time points (connected by line) are shown in plots. (C-D) Distribution of pre-existing subpopulation percentages of BCR-ABL1 V299L (C) and V299L compound (D) for those simulation results that fit our experimental observations. The histogram includes all parameter combinations with at least one simulation run (out of the 50 per parameter combination) that fit the data. The only way to explain our observed kinetics was with the pre-existence of BCR-ABL1 V299L. (E) Sensitivity analyses based on Monte Carlo sampling and stochastic birth/death model showing the effects of each parameter on final tumor population size and tumor composition, as measured by Kendall correlation. Blue and red indicate positive and negative correlation, respectively. The major determinant of final subpopulation sizes was the pre-existing subpopulation sizes. Birth/death rates and background mutation rates had
minimal effects. (F-G) Representative ODE simulation kinetics of long-term drug resistance selection with dose escalating concentrations of dasatinib. The dose schedule is shown in (F) and corresponding subpopulation fractions in (G). (H-I) Given the dose escalation simulations, we also predicted the EC50s for the overall population at each time point over the course of dasatinib selection. This provides an approximate treatment window for which we can observe temporal collateral sensitivity to drugs such as foretinib. The predicted EC50 for the overall population was based on a weighted sum of the known EC50s for individual subpopulations.
Figure 4. Sensitivity of BCR-ABL1 V299L acts through on-target inhibition of BCR-ABL1. (A).

Representative cell cycle profiles taken at 12 hours post treatment \textit{in vitro} for Ph+ ALL cell lines derived from drug selection experiments, either with BCR-ABL1 WT or V299L. Treatments with ABL1 inhibitors led to a G1 arrest, albeit at higher concentrations for V299L cell lines due to resistance. While no G1 arrest was observed upon treatment with crizotinib, foretinib, cabozantinib, and vandetanib in the...
BCR-ABL1 WT cell line, G1 arrest was observed in the presence of BCR-ABL1 V299L. (B) Representative flow cytometry analysis of phospho-Stat5 (a measure of ABL1 activity) and cleaved-PARP (a measure of apoptosis). Similar to the cell cycle profile phenotypes, we observed an inhibition of pStat5 in the presence of V299L upon treatment with the collaterally sensitives, supporting an on-target ABL1 inhibition as the mechanism of action. (C) In vitro kinase assay at 10 μM ATP with recombinant active ABL1 WT or V299L against vandetanib, showing a preferential inhibition of kinase assay against ABL1 V299L relative to WT. Results for other small molecules are shown in Supp Fig S8. Data are shown as mean ± s.d. from three independent experiments. (D-E) Models of vandetanib docked to ABL1 WT and V299L. Vandetanib and bosutinib are shown in blue and orange, respectively. V299L causes steric hindrance to nitrile group of bosutinib, whereas it provides additional van der Waals contact to quinazoline group of vandetanib.
Figure 5. Non-canonical BCR-ABL1 inhibitors demonstrates in vivo efficacy. (A) Representative in vivo bioluminescence of mice at and during time of treatment. Derived cell lines with either BCR-ABL1 WT or V299L was tail-vein injected into immunocompetent recipient mice. Initial imaging was performed at day 10 post transplantation. Mice were subsequently treated once daily with vehicle, 10 mg/kg dasatinib, 50 mg/kg imatinib, 50 mg/kg vandetanib, or 50 mg/kg foretinib. (B) Fold change in total whole-mouse bioluminescence signal between post and pre-treatment. Mice bearing BCR-ABL1 V299L ALLs showed significant tumor burden reduction upon treatment with foretinib or vandetanib. Statistical significance determined by Mann-Whitney test. * P < 0.05 and ** P < 0.01. (C-D) Spleen from the same cohort of mice was also imaged and weighted (4 days and 8 days post initial treatment for BCR-ABL1 WT and V299L, respectively). Treatment of BCR-ABL1 V299L in vivo with foretinib and vandetanib showed strong reduction in spleen size. Scale bar indicates 1 cm. Statistical significance determined by Mann-Whitney test. ** P < 0.01. (E) Kaplan-Meier overall survival of immunocompetent recipient mice transplanted with BCR-ABL1 WT or V299L. Treatment of mice with foretinib or vandetanib led to
significant extension in overall survival. Data presented was compiled from two independent injection experiments. Statistical significance determined with log-rank rest. *** $P < 0.001$. 
Figure 6. Sequential drug switching changes clonal trajectories. (A) Experimental setup of sequential drug selection. Parental ALL cell lines were initially selected with drug A at IC90 1x the concentration. Upon outgrowth, drug B was used at dose escalating concentrations for continued resistance selection. (B) Pharmacological profile depicting collateral resistance and/or sensitivity for recovered cells upon initial selection with drug A. Heatmap shows the log2 transform of the ratio in EC50s between given representative cell line with unique ABL1 mutation and the DMSO control cell line (similar as the parental). (C) Pharmacological profile for representative cell lines with initial unique ABL1 mutations following drug A → drug B selection. The sequence of drugs used for selection can diversify resulting resistant ABL1 mutations.
Figure 7. Small molecule screen reveals compounds with diverse fitness landscapes. (A) High-throughput small molecule screen with 391 compounds against parental, BCR-ABL1 V299L, BCR-ABL1 V299L/E255K derived ALL cell lines, as a model of the initial, intermediate, and terminal stages of clonal evolution. Plot shows the log2 ratio in EC50 between V299L and parental (for x-axis) and between V299L and V299L/E255K (for y-axis). Data points colored in orange are BCR-ABL1 inhibitor positive controls, and in red are other positive controls. (B-G) Conceptual fitness landscapes with predefined positions for the three clonal stages. Height of the peak/valleys determined based on actual EC50 values for given drug and cell line.
Figure S1. Evolution in fitness from stochastic branching process model, Related to Figure 1.

Evolution of parental population (as indicated by white asterisk in Fig 1A) upon drug selection. (A-B) A histogram as a function of time for population evolving toward higher fitness over 3000 generations. (A) Time evolution for drug A in the presence of continuous drug A selection. (B) Time evolution for drug D in the presence of continuous drug A selection, illustrating a treatment window with very low fitness if the population were to be exposed to drug D. (C-D) Simulation of selection using drug A for the first 2000 generations, followed by a switch to drug D. The fitness of resulting population over time for drug A (subpanel C) and drug D (subpanel D) is shown. Modeling results illustrate a diversification of trajectories and subsequent fitness using a sequential drug A \( \rightarrow \) drug D selection. Actual genotype of all subpopulations for simulations are shown in Supp Fig 2.
Figure S2. Genotype space of cell populations from stochastic branching process model, Related to Figure 1. Corresponding genotype space of population dynamics based on stochastic branching process model for Fig 1. Each point in the genotype space correspond to a genetically distinct population and can be high-dimensional in nature (i.e. variations in many genes). Here the genotype space is explicitly represented in a two-dimensional space. (A) Genotype space of cells over time under drug A selection. Cells evolved toward genotype associated with higher fitness (for drug A). (B) Genotype space of cells over time under drug A selection for first 2000 generations, after which switched to drug B selection. Cells evolved first toward greater fitness (for drug A), but upon drug switch exhibited different trajectories for evolution.
Figure S3. Variant calls from RNA-seq of parental and derived Ph+ ALL cell lines, Related to Figure 2. Varscan variant somatic mutation calls of independently derived cell lines at initial (IC90 1x) and terminal (IC90 16x) round of selection with dasatinib. Somatic calls were made using ‘normal’ as the parental and DMSO treated control cell lines. See methods for details of bioinformatics pipeline for generating the candidate hits. Resulting hits were annotated with SnpEff. (A-B) Bar plots of total number of unique somatic SNVs identified for initial round (A) and terminal round (B) of drug selection. Mutations that were categorized as missense/nonsense/UTR are colored blue, while others in grey. (C-D) Bar plots of total number of unique somatic indels identified for initial round (C) and terminal round (D) of drug selection. Mutations that were categorized as frameshift/UTR are colored blue, while others in grey.
Figure S4. Mutations in ABL1 domain from RNA-seq of parental and derived Ph+ ALL cell lines, Related to Figure 2. Short read alignments to ABL1 domain (with BWA) and mutations were visualized using Integrated Genomics Viewer (IGV). See Table S1 for outputs from VarScan somatic mutation caller.
Figure S5. Stochastic modeling robustly shows pre-existence of BCR-ABL1 V299L, Related to Figure 3. Distribution of pre-existing subpopulation percentages for parameter combinations where their simulation results matched experimental observations. Parameters (i.e. mutation rate, dose coefficient, initial subpopulation sizes, birth and death rates) were Monte Carlo sampled 10,000 times, and for each
given parameter combination 50 simulations runs were performed using a nonhomogeneous continuous-time birth-death stochastic model (see Methods). Specifically for birth/death values, Monte Carlo sampling was performed using a range of birth/death rate combinations at given experimental determined mean net growth rate for each subpopulation (A) or was further expanded to cover a range of net growth rates spanning their 90% confidence interval (B). Histogram shows the result based on any parameter sets with at least one simulation run fitting to experimental criteria assuming a 1% or 0.1% limit of detection (LOD). Resulting simulations were robust in all cases.
Figure S6 Mathematical modeling on effects of drug scheduling on time to relapse, Related to Figure 3. Expected time until relapse under different drug scheduling regimens and initial V299L subpopulation sizes at 0.1% (A) or 50% (B). Violin plots showing the distributions from simulation results of the time taken for the overall population at 2e6 cells to reach 1e14 (which is approximately the number of tumor cells close to terminal stage of tumor burden in our mouse model in vivo, based on our known kinetics of mice survival upon transplantation). Under the same total drug exposure, alternating drug schedule is superior over concurrent treatment when the initial V299L subpopulation fraction is small. Alternating regimen shown in pink, concurrent in green, and single drug in blue. Numbers in parentheses for concurrent regimens represent the ICx used. D, dasatinib. F, foretinib. Notation for ‘xd’ means each drug is dosed for x number of days before alternating.
Figure S7. Signaling measurements of Ph+ ALL parental and derived cell lines at high doses of cabozantinib, vandetanib, and imatinib, Related to Figure 4. Cleaved PARP and phospho-Stat5 measurements by flow cytometry of cells upon 12 h treatment with DMSO control, 1 μM or 10 μM of cabozantinib, vandetanib, or imatinib. Phospho-Stat5 was inhibited upon treatment with imatinib in all cell lines. In contrast, in all three independently derived cell lines containing BCR-ABL1 V299L, phospho-Stat5 was inhibited at the 1 μM of cabozantinib and vandetanib. This inhibition was persistent at the higher 10 μM dose.
IC50 (nM) | ABL1 WT | ABL1 V299L
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>std. error</td>
<td>mean</td>
</tr>
<tr>
<td>crizotinib</td>
<td>2606</td>
<td>790</td>
</tr>
<tr>
<td>foretinib</td>
<td>5640</td>
<td>866</td>
</tr>
<tr>
<td>vandetanib</td>
<td>1742</td>
<td>751</td>
</tr>
<tr>
<td>cabozantinib</td>
<td>&gt; 10,000</td>
<td>5340</td>
</tr>
</tbody>
</table>

Figure S8. Kinase assay shows selective inhibition of ABL1 V299L, Related to Figure 4. In vitro kinase assay at 10 μM ATP with recombinant active ABL1 WT or V299L against small molecules crizotinib, foretinib, cabozantinib, or vandetanib. These compounds revealed a preferential inhibition of kinase assay against ABL1 V299L relative to WT. Mean and standard error were calculated from three independent kinase assay experiments.
Figure S9. Docking of vandetanib to ABL1 V299L, Related to Figure 4. (A) Comparison of functional groups between bosutinib and vandetanib. Both molecules share groups that extend into the solvent region and occupy the adenine and hydrophobic pockets. (B) Surface model of ABL1 V299L (gray) in complex with docked vandetanib (blue).
Figure S10. Unsupervised hierarchical clustering of RNA-seq samples, Related to Figure 4.

Unsupervised hierarchical clustering of variance-stabilized RNA-seq data of parental and derived Ph+ ALL cell lines using DESeq in R (with Euclidean distance and complete linkage). All derived cell lines at initial round (IC90 1x) and terminal round (IC90 16x) dasatinib selection were clustered within their respective groups. The two parental cell lines and DMSO treated cell line passaged in parallel (during drug selection) were also clustered together.
Figure S11. RNA-seq differential expression analyses, Related to Figure 4. Differential expression analyses of RNA-seq data using DESeq in R, where dispersion was estimated based on a negative binomial distribution. Genes shown in red are considered differential expressed at a fold-change cut-off of 2 and adjusted p-value cutoff of 0.05. Subsequent gene set enrichment analyses of differentially expressed genes did not reveal any statistically significant functional categories.
SUPPLEMENTAL EXPERIMENTAL AND COMPUTATIONAL PROCEDURES

Proliferation assays
For dose responses, cells were plated in 96-well plates at 10,000 cells per well (50,000 cells/mL) with increasing concentrations with drug. After incubation for 48 hr, viable cells were counted on a flow cytometer (LSRFortessa or LSR II (BD Biosciences)) at a fixed volume and flow rate and gated with propidium iodide exclusion. Cell numbers were normalized to DMSO control. Resulting dose response curve were fit to the equation,

\[ y = A_{inf} + \frac{A_0 - A_{inf}}{1 + \left(\frac{\text{[drug]}}{EC50}\right)^n} \]

For proliferation assays, cells were plated in 96-well plates at 10,000 cells per well, and viable cell number were counted using flow cytometer at time points: 0h, 12h, 24h, 36h, and 48h. The net growth rate was determined from fit to an exponential growth equation: \( N = N_0 \exp(kt) \).

In vitro kinase assay
Kinase reaction was performed in 96-well plates in kinase buffer (Cell Signaling Technology; final 1X: 125 mM Tris-HCl (pH 7.5), 5 mM beta-glycerophosphate, 2 mM DTT, 0.1 mM NaVO₄, 10 mM MgCl₂) with 10 μM ATP, 0.2 μg/μL Abltide (SignalChem), purified recombinant kinase (1 ng active ABL1 V299L or 20 ng active ABL1 wild-type, from SignalChem), and drug at varying concentrations. The amount of kinase was determined through titration with standard curves to ensure that reaction at completion still remained within linear range and achieved 5-10% of ATP to ADP conversion. Reaction condition was also optimized to ensure at least a 10-fold signal over background. Reaction mix was incubated for 10 min prior to ATP addition. Kinase reaction was incubated at room temperature for 60 min, and amount of ADP was assessed using Promega ADP-Glo.

PCR amplification and Sanger sequencing ABL1 kinase domain
Genomic DNA was extracted with DNAeasy Blood & Tissue Kit (Qiagen). ABL1 kinase domain were PCR amplified using forward primer, 5'-TGTCTATGGTGTGTCCCCCA-3', and reverse primer, 5'-CCTGCAGCAAGGTACTCACA.-3'. The resultant 0.9 kb PCR product was Sanger sequenced with the same primers used for PCR, and additionally with forward primer, 5'-CTACGACAAGTGGGAGATGGAACG-3' and reverse primer, 5'-CCCAGCTCCTTTTCCACTTCGTCT-3'.

145
**RNA-seq alignments**

RNA was extracted using QIAshredder and RNeasy Kit (Qiagen). Quality of RNA was assessed with BioAnalyzer (Agilent Technologies) and library was prepared with Illumina TruSeq. Eight samples were sequenced on a single lane with 150 nt paired-end reads using NextSeq (Illumina). A total of 554.8 million paired reads were generated per lane. Sequence reads in FASTQ format were first trimmed to remove adapters and low quality reads. Processed reads were aligned using TopHat v2.0.12 \(^1\) to mouse genome/transcriptome (UCSC mm9), with the following settings (--no-novel-juncs --segment-length 20 --read-realign-edit-dist 0). For mutation calls specifically for BCR-ABL1, since the cDNA was transduced into murine cells, processed reads were aligned separately to ABL1 (NM_005157.5) using BWA v0.7.10 \(^2\).

**RNA-seq differential expression**

BAM files were sorted by read name and indexed using SAMtools v0.1.19 \(^3\). Reads were split into single and paired-end reads using SAMtools and counted using HTSeq v0.6.1 \(^4\). Counts were then combined and analyzed using DESeq R package \(^5\). Unsupervised hierarchical clustering was performed on variance-stabilized data and with Euclidean distance and complete linkage. Differential expression was performed with fold-change cut-off of 2 and adjusted p-value cut-off of 0.05. Pathway analyses were performed using DAVID v6.7 \(^6\).

**RNA-seq mutation calls**

BAM files were sorted by chromosomal coordinate and indexed using SAMtools v0.1.19 \(^3\). Mpileup files were generated using SAMtools with base alignment qualities turned off. Indels and SNVs were detected using VarScan v2.3.8 \(^7\) with somatic mutation calls (the ‘normal’ sample was labeled as either parental or DMSO treated control cell lines, containing BCR-ABL1 WT). High confidence hits were filtered with VarScan somaticFilter with the following settings, --min-coverage 10 --min-reads2 5 --min-var-freq 0.01 --p-value 0.05. SNV near indels were also filtered out. Resulting hits list were annotated with SnpEff v3.6 \(^8\). Final list of SNVs and indels for BCR-ABL1 V299L cell lines (selected at dasatinib IC90 1x) and BCR-ABL1 V299L compound mutant cell lines (selected at dasatinib IC90 16x) were somatic calls by combining lists compared against both parental and DMSO treated control cell lines. Short read alignments and mutations were visualized with Integrative Genomics Viewer (IGV) \(^9\). Final list of variants were compiled and analyzed in R.

*High-throughput small molecule screen*
Small molecule screen was performed as previously described (10) with modifications. Cells (derived cell lines with either BCR-ABL1 WT, V299L, or V299L/E255K) were seeded at 20,000 cells/mL in 384-well plates and incubated overnight. Next day compounds (Cambridge Cancer Collection, Selleck Chemicals) were pinned using Tecan Freedom Evo 150 (Tecan). Each drug was dosed at five concentrations with 10-fold dilutions in technical duplicates. After 48 h incubation, total cell number was assayed using Cell Titer-Glo (Promega) with luminescence signal read on the M1000 Infinite Pro plate reader (Tecan). Final screening data were normalized on a per-plate basis using the median of DMSO controls for each plate. Dose response data were fit first to a four-parameter fit (see equation under Proliferation assays) using minpack.lm package in R. Alternatively, if the solution did not converge and/or the quality of the fit was poor, a model with fixed A_{inf} and/or A_0 was used. All fits were manually inspected and any inactive compounds were labeled accordingly.

Cell cycle analyses
Cells were seeded at 0.8 million cells/mL and treated at the indicated drug concentration. At 12 h post-treatment, cells were pelleted and washed once with cold PBS/2% FBS and fixed with cold 70% ethanol (in PBS) at -20°C overnight. Cells were washed twice with PBS/2% FBS, suspended in staining solution (50 μg/mL propidium iodide (Sigma) and 200 μg/mL RNase A (Qiagen) in PBS), and incubated at 37°C for 15 min. Stained cells were analyzed on a flow cytometer (LSRFortessa or LSR II (BD Biosciences)).

Signaling measurements
Cells were seeded at 0.8 million cells/mL and treated at the indicated drug concentration. At 12 h post-treatment, cells were pelleted and washed once with cold PBS, and fixed in 4% formaldehyde in PBS at room temperature for 15 min. Fixed cells were washed once with cold PBS and suspended in 100% MeOH and stored overnight at -20°C. Next day cells were washed with 0.1% Tween-20 (in PBS) and stained with phospho-Stat5 (Y694) (Cell Signaling Technology, #9314), incubated for at least 1 h at room temperature. After incubation with primary antibody, cells were washed and incubated overnight with cleaved-PARP conjugated antibody (BD Pharmingen, #558710) and secondary Alexa Fluor 488 antibody (Invitrogen, #A-11008). Stained cells were analyzed on a flow cytometer (LSRFortessa or LSR II (BD Biosciences)).

In vivo studies
Two million parental or derived Ph+ ALL cell lines were injected into tail vein of 7 to 9 weeks old female syngeneic C57/BL6 mice (The Jackson Laboratory). At day 10 post-transplantation, mice were treated once daily for seven days via oral gavage with vehicle (80 mM citric acid, pH 3.1), 10 mg/kg dasatinib
(dissolved in 80 mM citric acid, pH 3.1), 50 mg/kg imatinib (dissolved in water), 50 mg/kg vandetanib (dissolved in saline, pH 4), or 50 mg/kg foretinib (saline, pH 4). For survival experiments, mice were monitored daily and were sacrificed once they became moribund. For imaging experiments, overall tumor burden in mice transplanted with luciferase-containing tumor cells were assessed with bioluminescence imaging using luciferin and Caliper IVIS imaging system, and analyzed using Living Image software v4.3.1 (PerkinElmer). Due to differences in tumor burden kinetics, mice with BCR-ABL1 WT were imaged on day 10 and day 13, and sacrificed on day 14, while mice with BCR-ABL1 V299L were imaged on day 10, 13, and 17, and were sacrificed on day 18. All mouse procedures were approved by Massachusetts Institute of Technology Institutional Animal Care and Use Committee.

Substrate specificity predictions using KINspect scores
BCR-ABL1 mutants found from our selection experiments were analyzed using the KINspect algorithm [cite], which was developed to predict critical residues driving substrate specificity. The algorithm aims to identify these determinants of specificity by exploring a large set of specificity masks (representing the importance of each residue for specificity - ranging from 0 to 1, with 1 being important for specificity, and 0 being not important) and identifying the best and most predictive mask. The resulting KINspect scores were used to evaluate the likelihood that the resistant mutants identified in this study would drive resistance by rewiring downstream signaling, with mutants hitting positions with low (or high) KINspect scores being the least (or the most) likely to lead to downstream rewiring.

Docking and energetics studies
Docking studies were performed using Schrödinger Suite 2015 (Maestro v9.8). We docked vandetanib independently to ABL1 WT and ABL1 V299L kinase domain active sites. The x-ray structure of bosutinib in complex with ABL1 (WT) kinase domain (PDB ID: 3UE4, 2.4 Å) was used because of the similarity in chemical structure between vandetanib and bosutinib. The protein structure was prepared with removal of bosutinib from the active site of the co-crystal structure and removal of crystal water molecules using the Protein Preparation Wizard. Ligand (vandetanib) conformation was generated using SMILES sequence, which was converted to a 3D structure using the LigPrep module (v45017). A 16 Å grid was generated centered on the active site. Docking simulation were performed with this grid using the Glide module (v63017). The ABL1 V299L mutant was prepared using the Maestro suite and the mutagenesis tool. A separate grid generation and docking simulation were performed for the V299L mutant. Visualizations were generated using PyMOL (Schrödinger).
Interaction energy calculations were performed on the docked structures using CHARMM molecular mechanics software \((11)\). If necessary, terminal side chain dihedrals were set to trans-planar (180 degrees) for residues HIS, ASN, and GLN. Hydrogens were added using the HBtIIId module with CHARMM22 force field. Missing side-chain atoms were also built using CHARMM. Ligand charges were determined using restricted electrostatic potential fitting (RESP) \((12)\).

**Stochastic branching process model (with predefined fitness landscape)**

We used a stochastic discrete-time branching process to model the evolutionary dynamics of cells on a predefined fitness landscape. The actual ruggedness of fitness landscapes remains unknown in most cases, and here we specifically defined a step-wise progression of fitness landscape consisting of two incremental peaks (for drug A) or a valley at the intermediate stage (for drug B) generated based on multivariate Gaussian distributions (with \(\mu = \{(12,12), (18,18), (24,24)\}\) and \(\sigma^2 = 20\)). Such landscapes based on Gaussian distributions have also been previously used in evolutionary dynamics simulations \((13)\). The dynamics were modeled using a stochastic branching process \((14)\). We initialized the simulation with a monomorphic parental population at low fitness (1e6 cells at coordinate (5,5)) (indicated by the white asterisk in the genotypic space in Fig 1) and ran for 3000 generations. At each generation, cells either die or divide with a probability associated to their fitness. As a result of assuming a Wright-Fisher process (as an approximation for the branching process, for computational efficiency), the cells divide to generate a random number of offspring based on a Poisson distribution. For each cell division, there is a probability that the cells mutate with a new fitness. The occurrence of mutation was modeled based on a binomial distribution, with probability equal to the mutation rate \(u = 10^{-6}\). The new genotype for mutated daughter cell was drawn from a multivariate Gaussian distribution \((\sigma^2 = 1)\) with the associated fitness as described according to the predefined fitness landscape. Model implementation and simulations were performed in Matlab on computing cluster with results analyzed using Matlab and Mathematica.

**Stochastic birth/death model**

We modeled our drug resistance selection experiments as a nonhomogeneous continuous-time multi-type birth-death process and a corresponding system of ordinary differential equations (ODE). We modeled three subpopulations consisting of the parental (with BCR-ABL1 WT), BCR-ABL1 V299L, and BCR-ABL1 V299L compounds (a representative V299L compound, V299L/E255K, was used). Each subpopulation has its own birth rate \(b\) and death rate \(d\). During each cell division, mutation can arise to generate BCR-ABL1 V299L from wild-type with a mutation rate \(\mu_1\); similarly BCR-ABL1 V299L
compounds from single V299L with a mutation rate $\mu_2$. Back mutations were neglected, by assuming an infinite-site model.

The propensity function $a(x)$ and state change vector $v$, as follows,

$$a(x) = \begin{cases} 
\frac{\mu_1 b_1 [\text{parental}]}{\Gamma} \\
\frac{d_1 [\text{parental}]}{\Gamma} \\
\frac{(1 - \mu_1)b_1 [\text{parental}]}{\Gamma} \\
\frac{\mu_2 b_2 [\text{V299L}]}{\Gamma} \\
\frac{d_2 [\text{V299L}]}{\Gamma} \\
\frac{(1 - \mu_2)b_2 [\text{V299L}]}{\Gamma} \\
\frac{d_3 [\text{V299L comp}]}{\Gamma} \\
\frac{b_3 [\text{V299L comp}]}{\Gamma} 
\end{cases}$$

$$v = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

where $\Gamma = (b_1 + d_1)[\text{parental}] + (b_2 + d_2)[\text{V299L}] + (b_3 + d_3)[\text{V299L comp.}]$. The propensity function describes the probabilities for the occurrence of each event per time step and the state vector describes the corresponding change to subpopulation sizes for given event.

The effects of drug were parameterized from experimentally determined dose response curves. We assumed that the cells grow exponentially in both the presence and absence of drug. In the absence of drug, the cell number (with initial population size $N_0$) grows with birth rate $b$ and death rate $d$. We used a parameter $\alpha$ to describe the effect drug on decreasing the net growth rate of the population. As such, the fraction of cells survived ($y$) can be described as,

$$y = \frac{N_0 \exp((b - d - \alpha)t)}{N_0 \exp((b - d)t)}$$

Solving for $\alpha$,

$$\alpha = -\frac{\ln y}{t}$$

Thus, this describes the effects of drug on growth rate given a known fractional survival $y$ measured at time point $t$. In practice, dose response curves were assayed at 48 h, and as such this was used for parameterizing the $\alpha$ parameter. Note that drug treatment can have its effect on birth and/or death rate.
Because we lacked additional quantitative measurements to fully characterize which or both parameters the drug would affect, we assumed here that the drug affects the death rate. In addition, we incorporated a dose coefficient ($\delta$) for the $\alpha$ parameter to account for additional variables such as time-inhomogeneity of the dose and medium evaporation. As such, the death rate induced by drug is $\delta \alpha$.

We parameterized the birth and death rates in the absence of drug using experimentally determined growth curves for each individual subpopulation. The growth curves provided the net growth rate of each subpopulation (0.06488 h$^{-1}$, 0.06259 h$^{-1}$, and 0.06293 h$^{-1}$ for parental, BCR-ABL1 V299L, and BCR-ABL1 V299L/E255K cell lines, respectively). In addition, the propidium iodide (PI) exclusion used during the proliferation assay also provided the fraction of viable and dead cell populations. We found that all cell lines had a reasonably stable viable cell fraction each day at ~95-99% PI-. As such, we estimated the death rate using the same equation derived for $\alpha$ previously.

We performed Monte Carlo sampling with a range of background mutation rates and initial subpopulation proportions. As recent studies suggest that the background mutation rate can be variable depending on covariates such as replication time, gene expression, and mutational categories, we sampled a broad range of mutational rates to account for additional uncertainties (15). We also sampled a range of birth/death rate combinations for the experimentally determined net growth rates for each subpopulation. The parameters and the corresponding range of values sampled are described below.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>Background mutation rate</td>
<td>$10^{-9}$ to $10^{-6}$ mut/bp/cell division</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>Background $\mu$ for V299L comp.</td>
<td>1 to 10 fold higher than $\mu_1$</td>
</tr>
<tr>
<td>$b_1$</td>
<td>Birth rate for parental</td>
<td>0.06488 to 0.06968 h$^{-1}$</td>
</tr>
<tr>
<td>$d_1$</td>
<td>Death rate for parental</td>
<td>0 to 0.0048 h$^{-1}$</td>
</tr>
<tr>
<td>$b_2$</td>
<td>Birth rate for V299L</td>
<td>0.06259 to 0.06719 h$^{-1}$</td>
</tr>
<tr>
<td>$d_2$</td>
<td>Death rate for V299L</td>
<td>0 to 0.0046 h$^{-1}$</td>
</tr>
<tr>
<td>$b_3$</td>
<td>Birth rate for V299L comp.</td>
<td>0.06293 to 0.06763 h$^{-1}$</td>
</tr>
<tr>
<td>$d_3$</td>
<td>Death rate for V299L comp.</td>
<td>0 to 0.0047 h$^{-1}$</td>
</tr>
<tr>
<td>$[V299L]_0$</td>
<td>Initial population size for V299L</td>
<td>0% to 1%</td>
</tr>
<tr>
<td>$[V299L\text{ comp.}]_0$</td>
<td>Initial population size for V299L comp.</td>
<td>0% to 1%</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Dose coefficient</td>
<td>1 to 3</td>
</tr>
<tr>
<td></td>
<td>Total initial population size</td>
<td>5 million</td>
</tr>
</tbody>
</table>
We performed sensitivity analyses with additional Monte Carlo sampling to also account for experimental uncertainties based on 90% confidence interval of the net growth rates. We sampled the same ranges of mutation rates and pre-existing subpopulations as before, and with the following range of birth/death rates: \( b_1(0.05520 \text{ to } 0.07456 \text{ h}^{-1}), \ d_1 (0 \text{ to } 0.0055 \text{ h}^{-1}), \ b_2 (0.05299 \text{ to } 0.07218 \text{ h}^{-1}), \ d_2 (0 \text{ to } 0.0053 \text{ h}^{-1}), \ b_3 (0.05415 \text{ to } 0.07170 \text{ h}^{-1}), \ d_3 (0 \text{ to } 0.0053 \text{ h}^{-1}) \). Our conclusions were robust to both the above and this broader sampling ranges (Supp Fig 6).

To estimate the percentage of pre-existing subpopulations, we constrained our simulation results to only those that fitted our experimental observations: at beginning of simulation, initial tumor composition consisted of 5 million cells with parental at greater than 99.9%, and BCR-ABL1 V299L and V299L compounds at less than 0.1%; at end of simulation at day 9, total population size at 30-50 million, with parental and BCR-ABL1 compound subpopulations both at less than 0.1% and BCR-ABL1 V299L at greater than 99.9%. We also constrained using 1% limit of detection instead of 0.1%, with no changes in our conclusions (Supp Fig 6).

The mathematical model was implemented using Matlab. In the exact stochastic simulation, the inter-event time is exponentially distributed with rate parameter equal to \( \Gamma \). In practice, for computing performance purposes, stochastic model was simulated using modified Poisson tau-leaping method, with error tolerance \( \epsilon \) set to 0.001 (16, 17). Monte Carlo sampling were done with 10,000 parameter combinations, and for each given combination, 50 simulation runs were performed.

**ODE model**

Our stochastic modeling results revealed that BCR-ABL1 V299L had to pre-exist at a sizable number (i.e. at \( \sim 0.0082\% \), this corresponds to 410 cells per 5 million total population size). We used a deterministic model with a system of ordinary differential equations (ODEs) to continue modeling our selection experiments with dose escalations, for computational efficiency. The model, as follows, is equivalent to our stochastic birth/death model in previous section.

\[
\frac{d[\text{parental}]}{dt} = b_1[\text{parental}] - d_1[\text{parental}] - \mu_1b_1[\text{parental}]
\]

\[
\frac{d[V299L]}{dt} = b_2[V299L] - d_2[V299L] + \mu_1b_1[\text{parental}] - \mu_2b_2[V299L]
\]
\[
\frac{d[V299L \text{ comp}]}{dt} = b_3[V299L \text{ comp}] - d_3[V299L \text{ comp}] + \mu_2 b_2[V299L]
\]

We performed simulations involving automatic dose escalations, whereby upon total population size reaching 50 million cells, the population was diluted to 5 million cells and drug concentration was escalated to twice the previous dose. Maximum drug concentration was bounded at 16x (to conform to what was done experimentally). Dilution was performed via random sampling based on a hypergeometric distribution. Monte Carlo sampling on the parameter ranges were performed similarly to that for stochastic modeling to account for the noise and uncertainty in our parameters. Mathematical models were simulated in Matlab, with ODEs solved using numerical solver ode45.

For drug scheduling simulations, the model was initialized with either a 0.1% V299L or 50% V299L subpopulation size (and rest as parental (BCR-ABL1 WT)). We performed Monte Carlo sampling over a range of background mutation rates and birth/death rates (same ranges as described above). The populations were exposed to dasatinib (at desired concentrations with ICx based on parental) and/or foretinib (at desired concentrations with ICx based on BCR-ABL1 V299L). Drug scheduling regimens explored consisted of alternating, concurrent, or single treatments.

REFERENCES

Chapter V

Concluding Remarks

Portions submitted (with modifications) as: Zhao, B., Hemann, M. T., & Lauffenburger, D. A. Modeling tumor clonal evolution for drug combinations design. *Under review.*
With the increasing focus on viewing cancer through an evolutionary lens, it will be advantageous – and perhaps even necessary – to equip the field with useful quantitative tools from population genetics, evolutionary dynamics, and engineering to understand how cancer evolves and respond to treatment. In this thesis, we presented several joint quantitative and experimental approaches to address how to rationally design drug combinations in light of intratumoral heterogeneity and tumor clonal evolution. In Chapter II, using a tractable model with precise control over initial tumor composition, we tried to address fundamentally how to design drug combinations in a heterogeneous tumor population. This has led to our findings, with in vitro and in vivo validations, that we can derive novel non-intuitive optimal drug combinations not containing the best component drug for individual subpopulations, as a consequence of combinatorial optimization. We extended this work in Chapter III to examine whether this conclusion hold for tumors with greater complexity. Most importantly, via statistical analyses, we found we can derive drug combinations even given some uncertainty in the initial heterogeneous tumor composition. In Chapter IV, under uncontrolled tumor clonal evolution, we examined more closely the natural trajectories of clonal trajectories under drug selection. Strikingly, we observed that predictable distinct stages of clonal trajectories can expose vulnerabilities (i.e. collateral sensitivity) for therapeutic intervention. As such, beyond optimization of drug combinations at a given time point, knowledge of tumor dynamics and trajectories in future time points can enable forward predictions and inform drug combinations design.

We conclude here with some thoughts and synthesis of emerging technologies and quantitative approaches that will shape our understanding of cancer biology. In particular, as alluded to in the Introduction chapter, several themes (Fig. 1) are emerging that address the critical questions of: 1) how can we predict evolutionary trajectories of tumor progression? and 2) how do we tackle drug resistance?

First, the advent of NGS will enable a much greater resolution of lineage tracking and at finer time scales in studying tumor clonal evolution. Although population genetic studies have benefited from evolution experiments and sequencing of lower organisms in yeast and bacteria, the utility of NGS specifically to study tumor evolutionary dynamics in vitro and in vivo at the desired level of resolution in regards to lineage complexity and time has been lacking. NGS efforts on studying patient biopsies have thus far provided greater understanding of the tumor clonal complexity and architecture. However, experiments need to be designed that capitalize on the availability of high complexity barcoding libraries and finer time resolution to specifically address questions on tumor evolutionary dynamics. These include the dynamics and distribution of beneficial, passenger, and deleterious mutations that arise (and go on to fixation or extinction) and their associated fitness effects. Mathematical modeling from population genetics is critical in parameterizing experimentally observed dynamics and in estimating parameters such
as mutation rate, selection coefficients, time to fixation, and regimes of dominance between genetic drift vs selection, etc. Additionally, an understanding of the clonal trajectories will also enable us to visualize (based on parameterized fitness landscape models and/or conceptually) the underlying fitness landscape (static or dynamic) as the tumor evolves over time. This will directly address our questions in regards to how rugged such landscapes are and how constrained are tumor evolutionary trajectories. As we have discussed in this review, knowledge of the fitness terrain would have direct implications on our understanding of repeatability and predictability of the tumor clonal evolutionary process.

Figure 1. Outlook of using enabling technologies and quantitative approaches to study cancer.
Cancer is currently viewed as an evolutionary process. This necessitates the design of empirical studies to specifically address cancer evolution as we have done in population genetics studies — including evolution experiments with high-resolution lineage tracing and mutagenesis/saturation analyses to systemically understand the fitness of specific alterations. Fitness landscapes provide one way to gain insight into how genotype or phenotype are mapped to fitness under different conditions. This information will be invaluable for the effective and rational design of optimal drug scheduling and drug combinations. Critical to these studies are the use of quantitative approaches rooted in population genetics, evolutionary biology, and engineering.
Viewing and understanding cancer through the evolutionary perspective also provides fresh opportunities for rational therapeutic intervention. As a particular fitness landscape can consist of peaks and valleys (and can be viewed as source and sinks in a vector field), this begs the question of whether we can control (or drive) the tumor toward particular desirable states. As we have discussed that some experimental models have started to address these questions, classical engineering approaches in control theory are one such approach to address the controllability of a given system. Alternatively, aside from control, our insight into the natural trajectories of tumor clonal evolution may reveal distinct stages along this path – beyond our current focus on terminal states – that are vulnerable for therapeutic intervention. These perturbations will no doubt have effects on the subsequent trajectories of tumor evolution. As such, exploring these trajectories using stochastic/deterministic mathematical models will be critical to explore these outcome scenarios. Furthermore, the search space for the choice and scheduling of drug to be utilized for these interventions are vast (including perhaps repurposing of drugs against non-classical targets), and optimization methodologies will facilitate the discovery of optimal solution (or solution sets) given a particular outcome scenario we desire. Taken together, we believe these new waves of enabling technologies and methodologies will be an important part of our toolset to advance our understanding in cancer research.