



Published in final edited form as:

Cell Rep. 2020 March 24; 30(12): 3951–3963.e4. doi:10.1016/j.celrep.2020.02.108.

## Multi-scale Predictions of Drug Resistance Epidemiology Identify Design Principles for Rational Drug Design

Scott M. Leighow<sup>1,4</sup>, Chuan Liu<sup>1,2,4</sup>, Haider Inam<sup>1</sup>, Boyang Zhao<sup>1</sup>, Justin R. Pritchard<sup>1,3,5,\*</sup>

<sup>1</sup>The Pennsylvania State University, Department of Biomedical Engineering, University Park, PA 16802, USA

<sup>2</sup>Department of Oncology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China

<sup>3</sup>The Huck Institute for the Life Sciences, Center for Resistance Evolution, The Pennsylvania State University, University Park, PA 16802, USA

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead Contact

### SUMMARY

Rationally designing drugs that last longer in the face of biological evolution is a critical objective of drug discovery. However, this goal is thwarted by the diversity and stochasticity of evolutionary trajectories that drive uncertainty in the clinic. Although biophysical models can qualitatively predict whether a mutation causes resistance, they cannot quantitatively predict the relative abundance of resistance mutations in patient populations. We present stochastic, first-principle models that are parameterized on a large *in vitro* dataset and that accurately predict the epidemiological abundance of resistance mutations across multiple leukemia clinical trials. The ability to forecast resistance variants requires an understanding of their underlying mutation biases. Beyond leukemia, a meta-analysis across prostate cancer, breast cancer, and gastrointestinal stromal tumors suggests that resistance evolution in the adjuvant setting is influenced by mutational bias. Our analysis establishes a principle for rational drug design: when evolution favors the most probable mutant, so should drug design.

### Graphical Abstract

---

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\*Correspondence: jrp94@psu.edu.

#### AUTHOR CONTRIBUTIONS

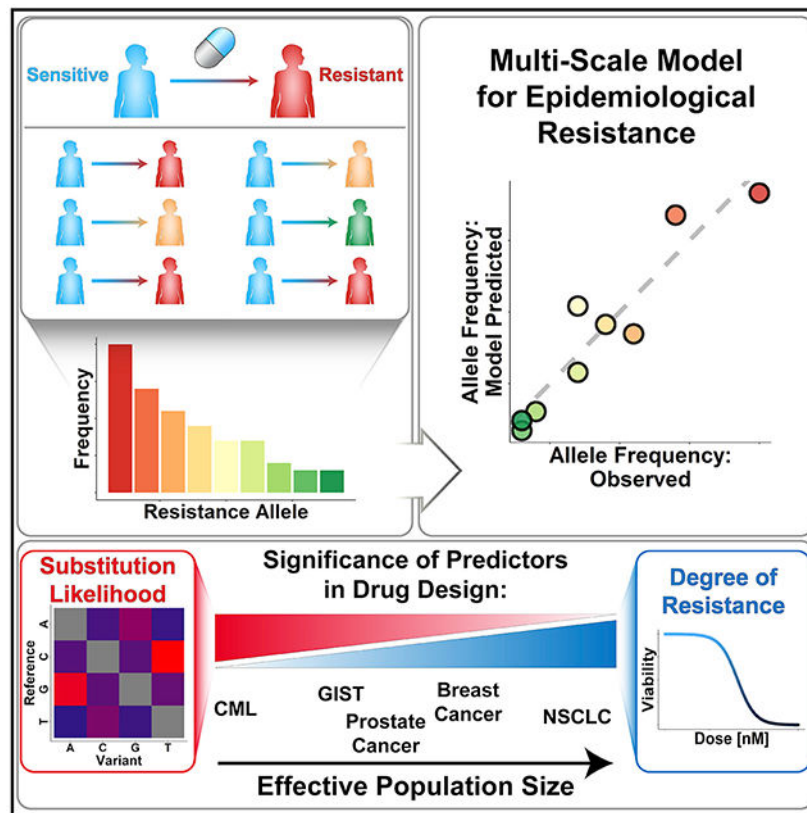
S.M.L. performed computational analysis and contributed to writing the paper. C.L. and H.I. performed experiments. B.Z. provided technical assistance. J.R.P. conceived the project, designed experiments, performed computational analyses, and wrote the paper.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

#### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.02.108>.



## In Brief

Drug resistance is often addressed through next-generation drug design, but evolutionary diversity complicates these efforts. Here, Leighow et al. demonstrate that multi-scale models can quantitatively predict mutant frequency. We find that when heterogeneity is limited, analysis requires an understanding of substitution likelihood. We show that these models can inform evolutionarily optimized drug design.

## INTRODUCTION

Since the time of Darwin, the most powerful demonstration of natural selection is the pervasiveness of genetic resistance following the adoption of new drugs for viruses, prokaryotes, eukaryotes, and cancers (Bonhoeffer et al., 1997a, 1997b; Goldberg et al., 2012; Hastings, 2004; Iwasa et al., 2006; Pritchard et al., 2012, 2013). Thus, efforts to rationally design new drugs that are less susceptible to evolutionary change are urgently needed.

Foundational stochastic models of evolutionary dynamics in cancer and infectious diseases have focused on the probability that most drug resistance mutations pre-occur in large populations of tumors, bacteria, and viruses (Bozic et al., 2013; Coldman and Goldie, 1983; Goldie and Coldman, 1984; Iwasa et al., 2006; Komarova and Wodarz, 2005; Luria and Delbrück, 1943; Michor et al., 2005). These theoretical arguments led to the practical insight that non-cross-resistant drug combinations are needed to combat genetic diversity. They also

formed the basis for our current therapeutic regimens in HIV, tuberculosis, and cancer. However, despite this powerful example of evolution-guided therapeutic regimen design, drug resistance remains a problem. We posit that an important step forward involves using decades of improvements in evolutionary theory (Coldman and Goldie, 1983, 1986; Goldie and Coldman, 1979, 1984; Iwasa et al., 2006; Komarova, 2006; Lässig et al., 2017; Lind et al., 2019; Lobkovsky and Koonin, 2012; Luka išinová and Bollenbach, 2017; Łuksza and Lässig, 2014) and population genetics (Gerrish and Lenski, 1998; LeClerc et al., 1996; Stoltzfus and McCandlish, 2017) to create additional design principles for drug discovery informed by evolution. We suggest that this might be achieved by expanding our ability to quantitatively predict which diverse resistance mutations can generate relapses in individual patients during treatment.

Recent and classic papers in cancer and bacteria have shown that biophysical methods and mutagenesis screens have great value in qualitatively identifying which mutations in a protein might lead to clinical resistance (Hall, 2004; Hauser et al., 2018; Rodrigues et al., 2016). However, a long list of possible resistance mutations is challenging to incorporate into drug design. Which mutations will be most clinically abundant and thus constitute must-hit variants during drug development? To answer this question, we must go beyond qualitative predictions of possible resistant mutant identity to quantitative predictions of evolutionary outcomes. Recent progress in predictive evolution has shown that biophysical metrics can predict cellular fitness landscapes in antibiotic resistance (Rodrigues et al., 2016) and that fitness effects can forecast seasonal flu trends (Lässig et al., 2017). These exciting steps forward led us to ask unanswered questions in drug resistance research: Can the diversity and distribution of mutations that contribute to resistance evolution be quantitatively predicted? And could those predictions guide drug discovery?

Two scales combine to quantitatively determine which drug resistance variants arise across a population: (1) the host-level variables affecting *de novo* resistance generation and (2) the community-level variables affecting the global spread of resistance. The *de novo* variables include growth rates in the presence and absence of drug (which are often measured), as well as mutation rate, codon structure, genetic context, and pharmacology (which are seldom measured). Human cancers offer a unique opportunity to investigate the process of parallel *de novo* resistance evolution in individual hosts, because they lack community-level variables that affect the spread of variants across a population (Figure 1A). Thus, predictive models of resistance to targeted cancer therapies are an interesting first test for evolutionary models of drug resistance variants across real-world populations.

The lack of a quantitative and predictive consideration of these evolutionary variables creates a gap in cancer patient care. In the targeted therapy of cancer, as first-generation drugs are used and resistance liabilities are identified, drug discovery scientists have raced to develop second-generation drugs that can target resistance mutations (Jabbour et al., 2015; Mok et al., 2017; O'Hare et al., 2009; Peters et al., 2017). Second-generation inhibitors can improve clinical outcomes in drug-resistant and drug-naive patients, but vulnerabilities persist because patient need drives rapid drug development in the face of immature clinical data (Jabbour et al., 2015; Mok et al., 2017; Peters et al., 2017). Thus, molecular design occurs before we know the true prevalence of specific mutations in the population, and

solutions are offered in the clinic before the full scope of the resistance problem is understood. Structure-based drug design is the industry standard to create potent second-generation inhibitors and has succeeded in ABL1+ leukemias (O'Hare et al., 2009; Wylie et al., 2017), c-KIT/PDGFR-mutated gastrointestinal stromal tumors (GISTs) (Evans et al., 2017), ALK/EGFR+lung cancers (Butterworth et al., 2017; Zhang et al., 2016), RET mutants/fusions (Subbiah et al., 2018), and TRK fusions (Menichincheri et al., 2016). Rational design is typically based upon the biophysics of binding to the target, not evolution. Using evolutionary theory to prospectively identify the residues and abundances that contribute to resistance following real-world drug use will improve pharmaceutical design (Hauser et al., 2018). By developing a broader picture of drug resistance evolution before clinical data have matured, evolutionary criteria may be combined with structure-function analysis to guide next-generation drug development.

In this study, we parameterize stochastic, first-principle models of drug resistance. By systematically studying multiple variables that could affect *de novo* resistance generation, we show that predictive evolutionary modeling can forecast population patterns of drug resistance without requiring clinical measurements of mutant-specific resistance parameters. Moreover, we show that multiple treatment scenarios and biological architectures create populations whose resistance evolution is sensitive to nucleotide substitution bias and codon usage. We posit that next-generation drug design could become more evolutionarily principled by adopting a simple design principle: when evolution favors the likeliest resistance mutation, so should drug discovery.

## RESULTS

### Analysis of a Single Salt Bridge in ABL1 Identifies Mutation Bias as a Correlate of Clinical Resistance Prevalence

We first examined two drug resistance mutations that exist at the same residue in the BCR-ABL oncogene: E255. E255 forms a salt bridge interaction with K247, stabilizing the phosphate binding loop (P loop) that clamps over the ATP competitive inhibitor in the active site of the molecule (Figure 1B). E255 can mutate to become either E255V or E255K, both of which abolish the charge interaction and promote clinical resistance to imatinib (Figure 1C), a BCR-ABL tyrosine kinase inhibitor (TKI) used to treat patients with chronic myelogenous leukemia (CML). Tallying the abundance of E255V and E255K mutations across clinical studies suggested that E255K was more prevalent (21 patients) than E255V (10 patients) ( $p < 0.05$ , chi-square test).

Next, we sought to examine why E255K might be more prevalent than E255V. We made BaF3 cells that harbored wild-type, E255V, and E255K BCR-ABL (a common model for BCR-ABL targeted therapies) (Shah et al., 2004), and we examined their relative sensitivity to imatinib. As expected, both mutations provided resistance to imatinib, but E255K provided significantly less resistance than E255V (Figure 1D). Importantly, this difference in phenotype occurs at clinically relevant concentrations of imatinib, which has a  $C_{\max}$  of 5.3  $\mu\text{M}$  and a  $C_{\text{trough}}$  of 2.4  $\mu\text{M}$  at steady state (Peng et al., 2004). Although further investigation into the biophysics of P loop dynamics could reveal the molecular mechanism behind this difference in drug resistance, we aimed to understand how a mutation that is worse at

conferring drug resistance might become more prevalent in humans. Because the less drug-resistant variant appeared in the clinic more often, we examined the relative growth rate of E255V and E255K in the absence of drug. If E255V grows significantly slower before imatinib exposure, that might help to explain its lower incidence in a human population. However, the relative growth rates for E255V, E255K, and wild-type BCR-ABL cells were statistically indistinguishable (growth rate  $\pm$  95% confidence interval [CI] of  $0.98 \pm 0.075$ ,  $1.03 \pm 0.125$ , and  $0.96 \pm 0.075$ , respectively; Figure 1E), despite having 80% power to detect a 10% growth rate difference. However, we do not rule out drug-free fitness effects on the basis of this measurement alone. We include drug-free growth rates for all mutants as a potential explanatory variable during model selection in Figure 3.

Thus, we hypothesized that a mutation from E > K might be more likely to occur than a mutation from E > V. Examining the genetic code, E > K requires a transition from G > A, whereas E > V requires a transversion from A > T. No other single-nucleotide substitutions can cause these two mutations. Although transitions are usually more likely than transversions, mutation biases are known to vary across genes and cancer types. We sought a direct measurement of the mutation bias in ABL1 and CML. To do this, we turned to multiple datasets that have analyzed mutations in cancer and variation in the normal human exome. As an example, the Broad ExAC dataset of 100,000 human exomes (see STAR Methods; Data S1) can be used to estimate ABL1 mutation bias by focusing on the extremely rare variants that constitute 90% of the variation in the ExAC data. Using this dataset, we examined rare (<1/10,000) mutations in the ABL1 gene (Figure 1F, top). Splitting these mutations by nucleotide type on the transcribed strand to account for biases because of transcription, we quantified the nucleotide substitution biases in the ABL1 gene. The measured bias was consistent with a transition-transversion bias; i.e., G > A mutations were more likely than A > T mutations. Moreover, it was largely invariant depending upon the exact mutation types analyzed (synonymous, nonsynonymous, and intronic) and consistent with previous literature. This ABL1 mutation bias was also highly correlated with the mutation biases that were measured across the CML exome (Figure 1F, bottom).

We observed that the substitution likelihood of a given amino acid mutation (defined as the nucleotide substitution bias and the individual codon path) is correlated with the clinical prevalence of E255V versus E255K.

### **An Analytical Model of Stochastic Dynamics Identifies Parameter Regimes in which the Likeliest Mutations Can Predominate**

Because E255K was more likely to occur than E255V, but not more fit in the presence or absence of drug treatment, we asked whether it makes theoretical sense that a more likely mutation can beat a more resistant one to create *de novo* drug resistance. To study this, we developed a probability model for the stochastic evolutionary dynamics of a hypothetical drug target with two possible resistance alleles. Allele A is more resistant (more fit in the presence of the drug) but less likely; allele B is less resistant but more likely (Figure 2A). The probability that either allele drives relapse was calculated from the allele-specific probabilities of seeding a detectable resistance clone. The probability that allele A is the dominant allele upon relapse is taken to be the sum of the probability that allele A is the

only mutant to spawn and the probability that allele A outcompetes allele B when both mutants spawn; the probability that allele B is the dominant allele upon relapse is similarly formulated.

The risk of mutation is modeled as a nonhomogeneous Poisson process in which the mutational hazard rate is a function of the size of the sensitive population responsible for seeding resistant clones. The overall probability of mutation to either resistance variant is given by the convergence of that allele's cumulative distribution function (Figure 2B, top). When both alleles are present, the race to outcompete the other allele considers the allele-specific fitness, as well as the time of mutation, as determined by the probability density function of mutational risk (Figure 2B, bottom). By considering the joint probability density function of mutation for the two alleles, it is possible to evaluate the probability that either allele drives relapse when both are present. From this model, we derived an asymptotic solution for the probability that either allele drives relapse. See STAR Methods and GitHub for complete descriptions.

We then evaluated these probabilities across a region of the effective population size/mutation rate parameter space to identify parameter regimes in which allele B (the less fit, more likely allele) is more likely to drive relapse. We refer to this phenomenon as survival of the likeliest, and it is strongest in the region of the phase plane with low mutation rates and a small effective population (Figure 2C), corresponding to populations with low heterogeneity. Here, where the opportunity for mutation is limited, unlikely mutants may not spawn, regardless of their fitness. Thus, under these conditions, it is more important for a resistance mutant to be likely than to be fit. In contrast, as mutation rate and effective population size increase, allele A is more likely to dominate, reflecting a more standard "survival of the fittest" model. Although this theoretical model was parameterized using values relevant to E255V/E255K, the results were qualitatively similar for other ABL1 variants (Figure S1). Thus, the degree of heterogeneity, as shaped by population size and mutation rate, determines when substitution likelihood can play a driving role in drug resistance epidemiology.

These theoretical results point toward the correlation of mutation and codon biases observed in real-world CML resistance incidence. The population structure of CML includes a well-characterized leukemic stem cell population of  $10^5$ – $10^6$  cells that gives rise to peripheral differentiated leukemic cells (Figure 2D). This hierarchy effectively restricts the population size, because only resistance mutations that occur in leukemic stem cells have any clinical consequence (Figure 2F).

Thus, CML can be placed in the low-heterogeneity regime of the parameter space. Altogether, theory and empirical data support the idea that low-heterogeneity populations are most strongly influenced by mutational likelihood.

### **Epidemiological Incidences of ABL1 Resistance Mutations Are Best Predicted by How Likely They Are**

To approach this problem more comprehensively, we gathered data from hundreds of parallel clinical evolution experiments in BCR-ABL+ leukemias across four continents over

17 years (Data S2; Bengió et al., 2011; Branford et al., 2003; Cortes et al., 2007; Hochhaus et al., 2013; Hughes et al., 2015; Kim et al., 2009). Specifically, we identified clinical trials with clear clinical resistance criteria and codon-level resolution. 268 high-confidence clinical cases of imatinib resistance were identified in these studies. Tallying mutations at individual amino acid positions, we found that the 19 most abundant amino acid mutations account for ~95% of the resistance mutations identified (Figure 3A). Notably, 85% of the resistance is caused by a mutation other than the fittest (T315I; imatinib  $IC_{50}$  = 8,711 nM, where  $IC_{50}$  is defined as the concentration to reach 50% of maximum inhibition).

For each of these 19 amino acids, we generated three independent isogenic BaF3 cell lines and measured the  $IC_{50}$  of imatinib in each cell line in triplicate across 11 serial drug dilutions (Figure 3B). Although this set of 19 clinical point mutations in ABL1 is the most systematically assembled *in vitro* dataset to date, many of these mutations have had their  $IC_{50}$  values measured in other resistance studies in other labs. Thus, the literature presents us with an opportunity to estimate how genetic context and experimental conditions might affect drug response. Figure S2A shows that cross-study correlations are high (Pearson's  $r$  = ~0.9 or higher for four studies), but systematic shifts exist between studies. This might suggest that the genetic drift in cell lines in an individual lab can influence drug potency and is consistent with recent findings (Ben-David et al., 2018). Thus, we have an opportunity to account for the effects of genetic background upon drug resistance. To do this, we normalized all systematic differences between the datasets into one mean dataset and combined all data from the literature with our data (see STAR Methods; Data S3). This cross-study approach leverages the genetic drift across cell lines to get the best estimate of the average isogenic effect of BCR-ABL mutations.

Beyond genetic-background-normalized  $IC_{50}$  measurements, we measured the drug-free growth rates of all ABL1 mutants from 6–10 independent lentiviral infections (Figure S2B). Many independent infections were critical, because the growth rates of individual clonal selections exhibit high variance. We also measured the substitution likelihood of a given amino acid conditioned upon a resistance mutation event using the frequencies from the ExAC analysis (Figure 3C; see STAR Methods and GitHub).

Negative binomial regression was used to predict the incidence of individual resistance mutants across the human population (Figure S2C; .rmd file on GitHub). We identified the best N-variable model by leave-one-out-cross-validation (LOOCV). By this metric, a two-variable model built on substitution likelihood and normalized  $IC_{50}$  values outperformed all others (Figure 3D). To verify the superior predictive power of these two variables, we also identified a likely independent dataset of different ABL1 mutation frequencies that is curated by the Sanger Institute. Although these mutations have a less clear clinical provenance than our analysis, we decided to use them as a test set, because they had considerably more data. Again, substitution likelihood and normalized  $IC_{50}$  outperformed all other combinations of predictor variables as evaluated by LOOCV (Figure 3E).

The two-variable model regressed on substitution likelihood and normalized  $IC_{50}$  values demonstrated remarkable accuracy, exhibiting a Pearson correlation between model-predicted and observed values of  $r = 0.76$  (Figure 3F). This model showed a marked

improvement over a two-variable model built on normalized  $IC_{50}$  values and drug-free growth rates ( $r = 0.27$ , Figure 3G). Although the statistical model in Figure 3F suggested that the degree of resistance and the substitution likelihood of a particular variant were both significant predictors of clinical abundance, substitution likelihood was the more significant ( $p = 3.3 \times 10^{-5}$  for substitution likelihood versus  $p = 0.016$  for  $IC_{50}$ , .rmd file on GitHub). To verify that our result was not overfit, we evaluated our statistical model against the Sanger Institute test set and found that the same two variables exhibited predictive accuracy (Pearson correlation  $r = 0.74$ , Figure S2D). However, in case our data have overlap with the Sanger Institute data, we also verified that the results were similar when our counts were subtracted from the Sanger counts ( $r = 0.72$ , Figure 3H). In addition, we considered the possibility that a fraction of the observations in our meta-analysis were included in the Sanger dataset. Therefore, we resampled the Sanger data by removing random subsets of our own data. Models built on these resampled data continued to exhibit high correlation between observed and model-predicted counts (Figure S2E). The reproducibility in a second dataset highlights the significance of substitution likelihood as the most important predictive variable.

Thus, substitution likelihood, rather than drug resistance, is the more significant variable predicting the relative abundance of the 19 mutants that account for 95% of the clinical mutations in BCR-ABL for patients with CML.

### **A Stochastic, First-Principle Model of Resistance Predicts the Clinical Prevalence of Resistance Mutations across ABL1**

Our experiments and epidemiological analysis indicated that an understanding of substitution likelihood is necessary to predict the clinically observed distribution of specific resistance mutations. Because of this, we wanted to know whether we could predict the frequency of mutations from a mechanistic model built from first principles.

We first identified a simple, clinically parameterized model of CML treatment that incorporates hematopoietic stem cell division and differentiation (Fassoni et al., 2018). To adapt the system to our question, we added our 19 resistance variants of interest to the model (Figure 4A). Each mutant was parameterized with an allele-specific substitution probability and a drug kill term. We used cell-based measurements in the presence of human serum to appropriately scale drug exposures to the effective *in vivo* levels (see STAR Methods).

Importantly, this methodology does not rely on fitting the model to the prevalence of specific resistance mutations. Therefore, it only requires a mechanistic model of treatment (i.e., birth/death rates), a list of candidate mutations (which could be generated from structure-based design or mutagenesis) (Figure S3A), measurements of substitution likelihood, and a tractable *in vitro* system to measure the effects of putative resistance mutations.

We simulated this system of 60 differential equations stochastically (Figure 4B) for 10,000 virtual CML patients treated with imatinib (see STAR Methods). Stochasticity allows the potential for random mutation to seed each resistant variant. Each patient was assigned a pharmacokinetic profile from clinically observed distributions of *in vivo* drug concentration (Peng et al., 2005). Similarly, patient-specific tumor detection sizes were drawn from real-



world distributions (Stein et al., 2011). Across these 10,000 simulations, we totaled the number of *in silico* patients who relapsed with each mutation (see STAR Methods).

Importantly, we ran simulations for a null model without substitution likelihood, in which individual amino acid changes were assigned the same probability, and compared its predictive value to that of a model parameterized with substitution likelihood. In examining the distribution of mutations across the ABL1 kinase, a model that considered substitution biases was required to accurately predict the abundances of individual mutations (Figures 4C and 4D;  $r = 0.21$  without versus  $r = 0.84$  with substitution likelihood).

To appraise the predictive power of the model for other drugs, we repeated this process of measuring  $IC_{50}$  values for all 19 resistance variants, simulating pharmacokinetic profiles, and parameterizing an evolutionary model for second-generation BCR-ABL inhibitors nilotinib and dasatinib. Both TKIs have been previously evaluated in frontline clinical trials (Hochhaus et al., 2013; Hughes et al., 2015). However, the count data for clinical resistance to nilotinib and dasatinib are sparse (0–3 patients for most variants) and suffer from wide CIs. Thus, to more completely evaluate our models' predictive power in the context of these two drugs, we examined categorical accuracy, i.e., the ability to say whether a mutation should or should not show up in a clinical trial. To do this, we downsampled our dasatinib and nilotinib simulations to reflect the size of the respective frontline trial. These virtual clinical trial results informed a binary classifier model—if a mutation was present, it was categorized as resistant; if it was not present, it was categorized as sensitive. We repeated this for  $10^3$  simulations, each time constructing a receiver operating characteristic (ROC) curve to quantify the accuracy of the binary classifier against the real-world clinical data. Averaging the ROC curves across all simulations (Figures 4E and 4F), we found that substitution likelihood considerably improved the models' ability to classify a variant as a resistance mutation (area under the curve AUC = 0.79 without versus 0.91 with mutation bias for dasatinib; AUC = 0.65 versus 0.82 for nilotinib). Given that these predictive models were not fit to patient data, they could be used to forecast which resistance mutations will be identified in a CML clinical trial *a priori*.

These results demonstrate that a first-principle, mechanistic model can predict the distribution of mutations seen in the clinic if it contains parameters that account for substitution likelihood. This improvement is consistent across imatinib, dasatinib, and nilotinib and underscores the importance of substitution likelihood in these models.

### **Evolution-Guided Drug Design Could Inform Principled Decisions between Mutational Vulnerabilities during Drug Development**

To further investigate how an evolutionarily informed approach to drug design might affect the clinical prevalence of resistance, we conceived of a hypothetical BCR-ABL TKI, which we call maxitinib, designed with mutational liabilities in mind. Given the same number of mutational vulnerabilities in a target (here we use five), maxitinib would be designed via structure-based drug design to target the five most likely mutants. If this could be achieved, how would that alter the overall incidence of resistance that arises in the clinic?

To test this, we simulated a cohort of *in silico* frontline CML patients for several target profiles of our hypothetical drug maxitinib; we name these distinct versions of maxitinib K1 to K15. Each of the hypothetical drugs, K1-K15 denote a version of maxitinib that was designed to target a different set of five of the 19 previously discussed imatinib-resistant mutants. The top five most likely mutants are sensitive to maxitinib K1, the second through sixth most likely mutants are sensitive to maxitinib K2, and so on (Figure 4G; see STAR Methods and GitHub for a model description).

The model predicts a 67% reduction in resistance incidence relative to imatinib for maxitinib K1, the hypothetical chemotype that targets the five most likely mutants (Figure 4H). This reduction is larger than either nilotinib or dasatinib (even though at least as many mutations confer resistance to maxitinib K1 as to nilotinib or dasatinib), and predicts that the evolutionarily optimal chemotype would prevail in the clinic. At the same time, maxitinib K15, which targets the five least likely mutants, offers virtually no advantage over imatinib. To be sure that the most prevalent resistance allele, T315I, was not dominating the observed effect, we removed T315I from the model and reperformed the simulations. Under these conditions, there was still high agreement between resistance incidence and mutational liability across the various chemotypes (Figure S3B).

Considerations relating to structure, conformation, and intermolecular interactions complicate drug design and would likely preclude the development of a drug that targets exactly the N most likely resistance mutants. However, as small molecules are screened and tradeoffs are identified, evolutionary models could choose a molecular profile that predicts less resistance when used in the clinic. Our models suggest that designing molecules in this way could create second-generation inhibitors that minimize drug resistance across a population (Figure S3C).

### **Mutation Biases Can Affect the Epidemiology of Resistance for Diseases Treated in the Adjuvant Setting**

Predictive modeling of resistance in pathologies beyond CML requires an understanding of what evolutionary forces determine the resistance allele prevalence in disease-specific settings. Although CML is a dramatic example in which substitution likelihood plays a dominant role in resistance evolution, a small-tumor-initiating population is not the only condition that restricts effective population sizes in cancer. In adjuvant therapy, a large tumor is surgically debulked before systemic treatment is initiated. Population bottlenecks created by surgical resection can considerably reduce the genetic diversity of tumor cells. We therefore hypothesized that mutation bias could shape the development of resistance to drugs used in the adjuvant setting.

Using a mechanistic model of tumor evolution, we simulated tumors treated with adjuvant therapy (see STAR Methods). In the model, a tumor grows (until  $10^9$ – $10^{12}$  cells, modeling a range of dissemination), at which point most of the tumor is excised and the remaining population ( $10^5$ – $10^8$  cells, modeling completeness of excision) is treated with an inhibitor. During division events, sensitive cells can spawn one of ten resistant alleles, each with randomly preassigned allele-specific resistance and mutational probability parameters. For each combination of tumor size pre- and post-surgery, 50,000 patients were simulated, and

their dominant resistance alleles were tallied upon relapse. The Spearman rank correlation of allele frequency with the degree of drug resistance (Figure 5A, top) and allele frequency with substitution bias (Figure 5A, bottom) were calculated for each combination of population size pre- and post-surgery. The simulation results show that when  $10^7$  or fewer cancer cells remained following resection, the allele frequency at relapse was more highly correlated with substitution bias than the degree of drug resistance conferred (Figure 5B).

To evaluate our theoretical results, we turned to clinical data for indications for which adjuvant therapy is often used. In particular, we analyzed on-target resistance mutations in the androgen receptor (AR) for prostate cancers with AR antagonists, in ESR1 for breast cancers treated with estrogen receptor (ER) antagonists, and in c-KIT for GISTs treated with imatinib (Figure 5C; Data S4). Some characteristics of the available clinical data complicate the analysis. Resistance data for these diseases were generally not delimited by whether the drug was administered adjuvantly, though we know based upon patient selection criteria that adjuvant patients were included in our cohorts. This ambiguity gives rise to a less pure, and likely more conservative, dataset, because our meta-analysis likely includes some patients with advanced metastatic disease. This uncertainty, combined with drug fitness data being unavailable for most resistance variants in the three drug targets, currently precludes an analysis identical to the one made in CML described in Figures 1, 2, 3, and 4.

We first noted that *in vitro* drug-sensitivity data are available in the literature for the most abundant resistance mutations for the three adjuvant diseases. The single-most clinically abundant allele was not the most drug resistant, contrary to what would be expected under a pure “survival of the fittest” model of evolution. Furthermore, there was poor correlation between the available  $IC_{50}$  data and the population allele frequency for these cancers (Figure S4A). This suggests that evolutionary forces beyond degree of resistance may shape the distribution of resistance alleles. We performed a similar analysis for on-target resistance mutations in ALK+ non-small-cell lung carcinoma (NSCLC) patients treated with crizotinib and in EGFR+ NSCLC patients treated with erlotinib or gefitinib. In NSCLC, targeted therapy is generally administered in the absence of surgical resection; thus, we would expect the larger effective population size to favor fitness over substitution likelihood (Figure 2F). For these two indications, the most frequent variant was also the most drug-resistant, and there was higher correlation between  $IC_{50}$  and clinical abundance (Figure S4B). Although these data are suggestive of complexity, they are not direct evidence that mutation bias influences resistance evolution in diseases treated in the adjuvant setting.

Given the constraints of the clinical data, we turned to an established technique for detecting mutational biases. The method has been used previously in cases of adaptive evolution within and across species (Payne et al., 2019; Stoltzfus and McCandlish, 2017). Specifically, the statistical test was developed to detect a simple case of mutation bias: transition-transversion bias. The approach involves aggregating observations of parallel evolution—in this case, drug resistance in cancer patients—and tallying the number of patients who develop each resistance mutation to determine the distribution of mutations (as in Figure S5A). Each mutation is classified according to the underlying nucleotide substitution type (transition or transversion), and the observed patient-level transition frequency is noted. Then, this distribution of mutations is simulated under the null assumption that mutation

bias does not influence evolution. That is, the dataset is simulated to reflect the expected outcome when variation in mutation frequency is decoupled from mutation bias. To do this, we reassign each mutation as a transition or a transversion according to its expected probabilities given no bias in mutation, and then we calculate the patient-level transition frequency. If the observed transition frequency is more extreme than those generated by the model, there is evidence that mutation bias influences evolution in the case at hand.

Although this statistical analysis is the literature standard for detecting mutational biases in parallel adaptive evolution (Payne et al., 2019; Stoltzfus and McCandlish, 2017), we noted that it had not previously been validated with simulated data. Therefore, we revisited our stochastic model of adjuvant therapy, for which we know that for small post-surgery population sizes, allele frequency is more correlated with mutation bias than with resistance. We considered the case with  $10^{10}$  cells pre-resection and  $10^6$  cells post-resection, and then we simulated these conditions with and without mutation bias. For the virtual cohort generated by simulations that included mutation bias, we found that the statistical test detected an overabundance of transitions ( $p = 0.001$ , Figures S5A-S5C). Conversely, the results of our control case simulated without mutation bias were not significant ( $p = 0.53$ , Figures S5D-S5F). Thus, our simulations confirmed the utility of the null model in detecting mutation bias, even with other confounding factors such as fitness.

We then turned to our meta-analysis for diseases treated with adjuvant therapy. The resampling technique indicated that the observed transition-transversion ratio (Figure 5D) exceeded those generated under the null model ( $p = 0.008$ , Figure 5E), suggesting that mutation bias influences the evolution of resistance in these diseases. We then expanded the resampling technique to develop a model of mutation bias that was more similar to our CML analysis by considering the six possible pyrimidine substitutions ( $C > A$ ,  $C > G$ ,  $C > T$ ,  $T > A$ ,  $T > C$ , and  $T > G$ , hereafter called mutation classes). We performed this similar analysis of simulating the resistance distribution by randomly reassigning the mutation class associated with each variant and tallying the number of patients in each class (simulated data in Figures S5G-S5N). We then calculated the Mahalanobis distance from the mean for both the simulated and the observed data (Figure S5O) and found that the observed data were more extreme than those simulated under the null model ( $p = 0.045$ , Figure 5F). See STAR Methods for more details.

This effect, although significant, is still not as strong as in ABL1+ CML ( $p = 0.018$ , Figure S5P). As previously noted, the data gathered across clinical trials likely include patients who did not receive adjuvant care, diluting the effect of restricted population sizes. Furthermore, even for those tumors that were treated in the adjuvant setting, it is likely that the effective population size and mutation rates in these diseases place them closer to the border of the phase change in Figure 2F. Given the understanding that mutation bias is likely to play a somewhat weaker role in resistance evolution for adjuvant diseases, we had aggregated the observations from prostate cancer, breast cancer, and GISTs to maximize our power to detect an effect. Prior studies that use this resampling technique also aggregated their data (Stoltzfus and McCandlish, 2017), combining observations across many taxa, mutation rates, and putative fitness effects. Even so, we wanted to confirm that our rejection of the null hypothesis was not an artifact of a large, aggregated dataset. To address this, we considered

the distribution of resistance mutations in ALK+ and EGFR+ NSCLC, two nonadjuvant indications, as well. The aggregated NSCLC data are not significant on their own ( $p = 0.45$ ). Furthermore, combining the NSCLC counts to our dataset for adjuvant diseases does not simply make the results more significant ( $p = 0.44$  for all diseases versus  $p = 0.045$  for adjuvant diseases alone, Figure 5G). This suggests that for the adjuvant diseases, the observed significance was not an artifact of aggregated data and increased sample size. This result is supported by a forward selection method used to determine the combinations of disease that show significant evidence of mutation bias. Such an approach identifies the most significant aggregated dataset as the one that includes the adjuvant diseases and excludes the nonadjuvant diseases (Figure S5Q; see STAR Methods for details).

These results indicate that the approach outlined here is a sensitive and specific method to identify cases in which substitution likelihood shapes drug resistance evolution. Such cases represent indications for which next-generation drug design could be informed by analysis of mutation biases and substitution likelihood.

## DISCUSSION

In this study, we demonstrate the ability to accurately predict real-world resistance evolution with amino-acid-scale resolution. In cases of restricted genetic diversity, such as in populations with small effective sizes, this predictive power requires an understanding of the likelihood of substitutions that generate the variation upon which selection acts. To be clear, it is inaccurate to say that drug selection is unimportant. A variant that grows out during treatment must still harbor a drug resistance phenotype, and our analysis of CML epidemiological data (Figure 3) indicates that the degree of resistance is still a significant predictor of clinical prevalence. However, in CML and other mutation-limited diseases, the ability to forecast amino acid prevalence requires a quantitative understanding of the stochastic molecular events that underlie resistance evolution. Beyond CML, our theoretical and empirical analyses (Figure 5) indicate that drugs used in the adjuvant setting provide multiple additional examples in which resistance evolution is influenced by substitution likelihood.

Beyond the retrospective analyses of molecular evolution, our study highlights the ability to prospectively forecast macroscopic evolutionary outcomes. Recent studies have shown the power of multi-scale modeling in forecasting the evolutionary trajectories of influenza evolving immune escape (Łuksza and Lässig, 2014) and predicting fitness landscapes of bacteria evolving resistance against trimethoprim (Rodrigues et al., 2016). Our work builds on these findings by predicting epidemiology-scale resistance from molecular insights. Although *in vitro* parallel evolution experiments have been able to qualitatively nominate resistance variants (Figure S4A), we demonstrate the ability to quantitatively predict interpatient heterogeneity with a completely mechanistic model of intratumoral heterogeneity.

Despite their power, our predictions are an imperfect step forward. Although the accuracy of these models is surprisingly high given their simplicity and that none of the mechanistic parameters were fit to clinical data, there are other variables that could explain the small

amount of residual variance. The first is that the genetic background outside of ABL1 can alter the level of phenotypic drug resistance. Furthermore, although fitness in the absence of drug was not a predictor of epidemiological abundance in our statistical models, it is still possible that drug-free fitness *in vivo* is distinct from our measurements *in vitro*. Despite these caveats, our work fits most variance in the data. This highlights the promise of evolutionary forecasting and emphasizes how predictive models enable a deeper understanding of evolution.

The basic concept of evolutionarily informed treatment regimens that minimize drug resistance is not new. Theoretical evolutionary models (Bozic et al., 2013; Goldie and Coldman, 1984) support empirical clinical evidence (Frei, 1985) that non-cross-resistant combination therapy can result in durable responses for some patients. However, effective non-cross-resistant drug combinations are not available for many cancers, and multiple drugs can cause overlapping off-target toxicities. In this study, we introduce a quantitative evolutionary design principle for single-agent therapy. When effective population sizes are small, evolution favors the most likely resistance mutation, and so should drug discovery (Figures 4G and 4H). This is not simply a matter of trying to reduce the absolute number of mutational paths to drug resistance. Prospective therapeutic leads can be prioritized for evolutionary optimality. Importantly, favoring either substitution likelihood or fitness alone would only be useful in isolation in extreme cases. There is a continuous spectrum of relative importance for both fitness and mutation bias as a disease transitions from low to high mutational complexity parameter sets. In all cases, we believe that rational drug design may be synonymous with evolutionarily informed drug design. Predictive evolutionary models like the ones presented here will enable precision medicines that are more efficacious and longer lasting in the face of biological change.

## STAR★METHODS

### LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Justin Pritchard (jrp94@psu.edu).

All unique/stable reagents generated in this study (including BCR-ABL BaF3 cell lines) are available from the Lead Contact with a completed Materials Transfer Agreement.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

**BaF3 Cells**—BaF3 cells were ordered from DSMZ (catalog number: ACC-300). BaF3 cells are a male mouse pro B cell line. BaF3 cells are maintained in 1640 (Sigma Aldrich) +10%FBS(Fisher)+1%Penn/Strep (Life Technologies) and 10ng/mL IL-3 (PeproTech).

**CML Clinical Data**—Clinical resistance data used in our meta-analysis was identified from six studies of frontline imatinib use with clear clinical resistance criteria and codon level resolution (Data S2; Bengi6 et al., 2011; Branford et al., 2003; Cortes et al., 2007; Hochhaus et al., 2013; Hughes et al., 2015; Kim et al., 2009). Across the six studies, we identified 268 cases of imatinib resistance. We similarly collected ABL1 resistance mutations reported in frontline clinical trials of second-generation inhibitors nilotinib (27

resistance cases; Hochhaus et al., 2013) and dasatinib (14 resistance cases; Hughes et al., 2015). Clinical resistance data for CML was also identified from a dataset curated by the Wellcome Trust (1322 resistance cases). The Sanger Wellcome Trust download was as of 12-01-2017. (<https://cancer.sanger.ac.uk/cosmic/csamples/details>).

**Clinical Data Used in Adjuvant Therapy Analysis**—Clinical data of high confidence resistance cases was collected from trials with clear criteria for resistance and amino acid resolution of target mutations for three diseases often treated in the adjuvant setting (AR+ prostate cancer, ER+ breast cancer, and KIT+ GIST) and two diseases generally treated with an inhibitor alone (EGFR+ and ALK+ NSCLC). Mutation counts for each disease are found in Data S4. In cases where the reported amino acid change was insufficient to infer the base pair mutation and nucleotide-level mutations were not reported, the count was omitted from mutation class analysis.

## METHOD DETAILS

**Construct Generation**—Following recombination-based cloning into pLVX-IRES-Puro, site directed mutagenesis was utilized to make the correct mutation in BCR-ABL. Mutation identity was confirmed by Sanger sequencing.

**Lentiviral Transduction**—Lentiviral constructs were co-transfected in HEK293T cells (ATCC catalog number CRL-3216) using calcium phosphate alongside third-generation packaging vectors that were pseudotyped with VSV-G. Viral supernatant was collected at 24 hours (Gozgit et al., 2018). All BCR-ABL mutations were infected at limiting MOI to achieve the lowest viral titer required to produce IL-3 independence. After selection in the absence of IL-3, we tested for puromycin resistance. All engineered cell lines were sequenced in the BCR-ABL kinase domain to confirm their identity.

**IC<sub>50</sub> Measurements**—Eleven serial dilutions were performed. Imatinib, dasatinib and nilotinib were all obtained from Selleck Chem. Starting points for dilutions were 10uM, 100nM, and 1uM respectively. 3000 BaF3 cells with the indicated mutation were seeded into a 96 well plate in 150ul of RPMI 1640 (Sigma Aldrich) +10%FBS (Fisher)+1%Penn/Strep (Life technologies). After addition of the drugs, cells were left in the incubator for 72 hours. At 72 hours Cell Titer Glo (Promega) was added at 1:4. This is less than the manufacturer's instructions, because we have verified the sensitivity of the assay with this reduced protocol. Plates were read via a luminescence plate reader at 72 hours.

**Serum Protein Shifts**—HSA-AAG containing medium is RPMI 1640 growth medium with the addition of 341 mM HSA (human serum albumin, Sigma, Cat # A9511) and 1 mg/ml AAG (human a1-acid glycoprotein, Sigma, Cat # G9885). All medium is sterilized by filtration through a 0.22mm membrane. Following the formulation of HSA-AAG media, identical IC<sub>50</sub> curve experiments are performed. The fold change of the IC<sub>50</sub> across 3 mutants of varying affinities (M244V, E255V, and WT BCR-ABL were used). Serum protein binding decreases the free drug concentration and shifts IC<sub>50</sub> values to numbers that are directly comparable to measured *in vivo* pharmaceutical exposures.

**Exome Data Analysis**—Data from the broad ExAC consortium (<http://exac.broadinstitute.org/>) for ABL1 was downloaded as of 11-17-2017. Raw data and processing code are available in Figure 3 in our GitHub repository: <https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure3/Mutation%20probabilities>

We tallied each of the 12 possible nucleic acid substitutions (not six) because transcription coupled repair has been shown to cause biases in the mutational spectrum of the transcribed strand, and ABL1 is widely expressed. We also compared the distribution of substitutions across ABL1 in the ExAC data to genome-wide mutation biases measured in CML exomes, as well as simple transition/transversion biases from the literature, and we did not observe significant differences.

**Theoretical Model of Competing Resistance Alleles**—We developed a probability model of two competing resistance alleles. The model considers a population of sensitive cells that expand and then contract under drug treatment. During the lifetime of the sensitive cells, a division event may result in mutation to one of the two resistance alleles, each with allele-specific drug growth rates and mutational probabilities that govern their population dynamics. We are interested in determining the probability that either allele is detected first, because we are interested in explaining the clinical data and the first mutant to detection is generally the one reported in a frontline clinical trial. We consider the probability that allele  $i$  is detected first to be the sum of the probability that allele  $i$  is the only one to spawn and the probability that allele  $i$  reaches detection first when both alleles spawn. We model mutation as a nonhomogeneous Poisson process (dependent on the size of the sensitive population responsible for seeding resistance); we model outgrowth to detection as an exponential process. These two quantities determine the probabilities of when and if either allele reaches detection. From these probabilities we derive an asymptotic solution for the probability that either allele is dominant upon relapse. For a complete discussion of the analysis, see the Appendix S1 LaTeX file on GitHub: <https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/SuppAppendices/SuppAppendix1>

**CML Stochastic Model**—We developed a stochastic model of evolutionary dynamics in chronic myelogenous leukemia. Specifically, we expanded the clinically-parameterized model in Fassoni et al. (2018) to include the 19 resistance mutants analyzed in this study. The mutants were parameterized with allele-specific mutational probabilities and degrees of resistance. Mutational probabilities were measured from the ExAC data as described. To appropriately parameterize drug kill rates, we simulated pharmacokinetic profiles using distributions of clinical exposure reported in Peng et al. (2004) and translated drug exposure to drug killing using the serum-shifted dose response curves. Another significant model parameter, tumor burden upon initial detection, was drawn from the distribution reported in Stein et al. (2011). The model was simulated for 10,000 virtual patients using MATLAB. For a complete description of the model, see the Appendix S2 LaTeX file on GitHub: <https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/SuppAppendices/SuppAppendix2>



**Adjuvant Therapy Stochastic Model**—We developed a stochastic model of resistance evolution for a theoretical tumor treated with adjuvant therapy. To do this, we modeled a small, sensitive initial population that expanded until detection at a population size of  $M_{pre}$  cells. Upon detection, the tumor is surgically debulked (reduced to a population size of  $M_{post}$  cells) and drug treatment is initiated, under which the sensitive population has a negative growth rate. During the lifetime of the sensitive population, division of sensitive cells could spawn resistance mutants. A resistance mutation could give rise to one of ten alleles, each with randomly pre-assigned allele-specific substitution likelihoods and degrees of resistance. These randomly assigned parameters were identical across all simulations. Upon relapse (defined as disease progression to initial detection size), the dominant allele upon relapse is tallied. This system was simulated stochastically for a range of  $M_{pre}$  and  $M_{post}$  values, and for each combination 50,000 virtual patients were analyzed using MATLAB. For a complete description of the model, see the Appendix S3 LaTeX file on GitHub: <https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/SuppAppendices/SuppAppendix3>

## QUANTIFICATION AND STATISTICAL ANALYSIS

**BCR-ABL BaF3 Fitness Analysis**—The drug-free growth rates and drug-dose response for BCR-ABL BaF3 variant was measured. For each variant,  $N = 3$  independent infections were conducted. For drug-dose response experiments, 11 serial dilutions of drug were dosed in triplicate for each independent infection of all variants. For more information, see the Figure 3 legend. For drug-free growth rate experiments,  $N = 11-14$  biological replicates were used due to the high fitness variability for low-titer infections. Error bars in figures are mean  $\pm$  standard deviation. Growth rates across variants were compared using Student's t test. For more information, see the Figure 1 legend.

**CML Epidemiological Analysis**—Negative binomial regression was used to construct models for resistance count data. The best model (using our meta-analysis and the Sanger dataset) was identified using LOOCV. Variable significance is calculated using R's glm function for negative binomial regression (Wald's chi-square test). For a complete analysis, see supplemental R markdown files and code repository for Figure 3. Code was written in R markdown. Raw data and analysis code (.Rmd and knitted .html files) are provided on GitHub: <https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure3>

**Clinical Adjuvant Therapy Analysis**—We compiled data for resistance variants in AR, ESR1, and c-KIT for prostate cancer, breast cancer, and GIST, respectively, and classified each count by the associated mutation type (transition or transversion) and pyrimidine substitution ( $C > A$ ,  $C > G$ ,  $C > T$ ,  $T > A$ ,  $T > C$ , and  $T > G$ ; mutation class). For cases when an amino substitution could be caused by multiple single nucleotide mutation paths, we classified them when nucleotide substitution resolution was available in the literature and excluded the counts otherwise. Under the null model, these variants were randomly reassigned mutation classes and the sum of all counts for each mutation class was noted. Since our data only considers mutations that result in an amino acid change, the mutation class assignments were weighted by the number of mutations in each class that are

nonsynonymous (116 transitions and 276 transversions; 68 C > A, 76 C > G, 58 C > T, 64 T > A, 59 T > C, 68 T > G). 10,000 replicates were simulated for each dataset. The Mahalanobis distance of each simulation from the mean of all simulations was calculated; the Mahalanobis distance of the observed mutation class counts from the simulation mean was also determined. An empirical p value was obtained as  $p = r/n$ , where n is the number of simulations and r is the number of simulations with a Mahalanobis distance greater than or equal to that of the observed data. Code was written in R markdown. Raw data and processing code are available in our GitHub repository: <https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure5>

**Forward Selection of Diseases for Analysis**—In order to verify that aggregating counts from adjuvant diseases was not artificially creating significance, we evaluated the significance yielded by several combinations of diseases, both adjuvant (AR+ prostate cancer, ER+ breast cancer, and KIT+ GIST) and non-adjuvant (ALK+ and EGFR+ NSCLC). To do so, we considered a stepwise approach analogous to forward feature selection in regression. We began by analyzing each disease individually to identify the one with the most significant mutational signature. We then appended data from the remaining diseases individually to determine the most significant paired dataset. We repeated this to identify the most significant K-disease dataset for  $K = 1$  to  $K = 5$ . This forward selection technique identified the most significant combination to be the 3-disease dataset that included our adjuvant diseases (prostate cancer, breast cancer, and GIST). See Figure S5R for results. Code was written in R markdown.

## DATA AND CODE AVAILABILITY

The datasets and code generated during this study are available at GitHub. Specifically, results from the BCR-ABL BaF3 experiments (including data on dose response and drug free growth rates) can be found at <https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure3>. Code used to execute the CML stochastic simulations can be found at <https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure4>. Code used for the adjuvant therapy model and code used to detect mutation bias in allele distributions can be found at <https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure5>.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

The authors thank Vivek Kapur, Mike Schmitt, Victor Rivera, Mengrou Lu, Kyle McIlroy, Lauren Randolph, Maciej Boni, Jeremy Rock, Igor Aronson, and David Kennedy for their contributions. Attendees at the 2018 Gordon conference on drug resistance were also critical to shaping this manuscript. This work was supported by an R21 from the NIBIB (5R21EB026617-02).

## REFERENCES

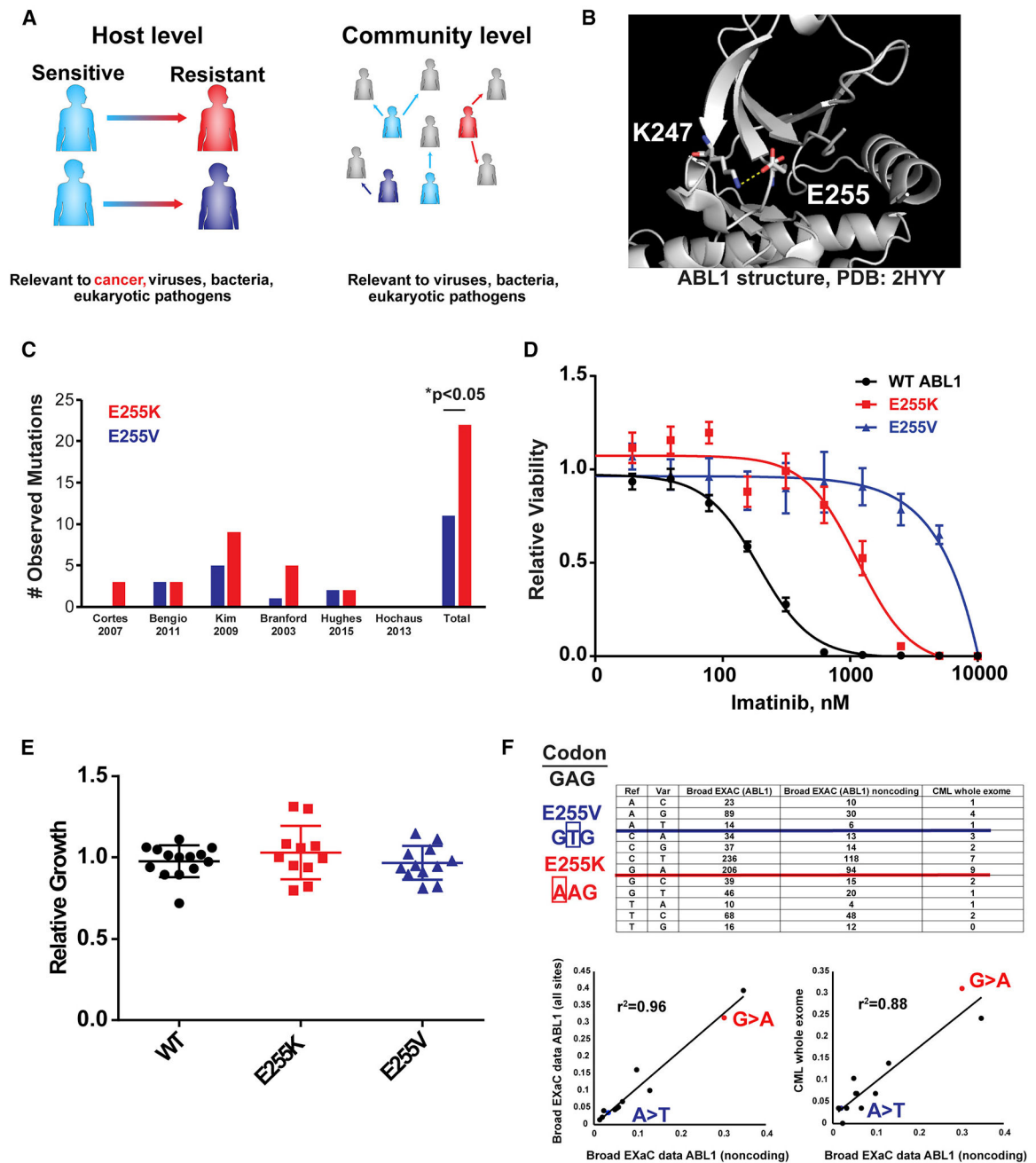
- Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, Strathdee CA, Dempster J, Lyons NJ, Burns R, et al. (2018). Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* 560, 325–330. [PubMed: 30089904]
- Bengi6 RM, Riva ME, Moiraghi B, Lanari E, Milone J, Ventriglia V, Bullorsky E, Tezanos Pinto M.d., Murro H, Bianchini M, and Larripa I (2011). Clinical outcome of chronic myeloid leukemia imatinib-resistant patients: do BCR-ABL kinase domain mutations affect patient survival? First multicenter Argentinean study. *Leuk. Lymphoma* 52, 1720–1726. [PubMed: 21663510]
- Bonhoeffer S, May RM, Shaw GM, and Nowak MA (1997a). Virus dynamics and drug therapy. *Proc. Natl. Acad. Sci* 94, 6971–6976. [PubMed: 9192676]
- Bonhoeffer S, Lipsitch M, and Levin BR (1997b). Evaluating treatment protocols to prevent antibiotic resistance. *Proc. Natl. Acad. Sci* 94, 12106–12111. [PubMed: 9342370]
- Bozic I, Reiter JG, Allen B, Antal T, Chatterjee K, Shah P, Moon YS, Yaqubie A, Kelly N, Le DT, et al. (2013). Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife* 2, e00747. [PubMed: 23805382]
- Branford S, Rudzki Z, Walsh S, Parkinson I, Grigg A, Szer J, Taylor K, Herrmann R, Seymour JF, Arthur C, et al. (2003). 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. [PubMed: 12623848]
- Butterworth S, Cross DAE, Finlay MRV, Ward RA, and Waring MJ (2017). The structure-guided discovery of osimertinib: the first U.S. FDA approved mutant selective inhibitor of EGFR T790M. *MedChemComm* 8, 820–822. [PubMed: 30108799]
- Coldman AJ, and Goldie JH (1983). A model for the resistance of tumor cells to cancer chemotherapeutic agents. *Math. Biosci* 65, 291–307.
- Coldman AJ, and Goldie JH (1986). A stochastic model for the origin and treatment of tumors containing drug-resistant cells. *Bull. Math. Biol* 48, 279–292. [PubMed: 3828558]
- Cortes J, Jabbour E, Kantarjian H, Yin CC, Shan J, O'Brien S, Garcia-Manero G, Giles F, Breeden M, Reeves N, et al. (2007). Dynamics of BCR-ABL kinase domain mutations in chronic myeloid leukemia after sequential treatment with multiple tyrosine kinase inhibitors. *Blood* 110, 4005–4011. [PubMed: 17785585]
- Evans EK, Gardino AK, Kim JL, Hodous BL, Shutes A, Davis A, Zhu XJ, Schmidt-Kittler O, Wilson D, Wilson K, et al. (2017). A precision therapy against cancers driven by KIT/PDGFRA mutations. *Sci. Transl. Med* 9, ea01690. [PubMed: 29093181]
- Fassoni AC, Baldow C, Roeder I, and Glauche I (2018). Reduced tyrosine kinase inhibitor dose is predicted to be as effective as standard dose in chronic myeloid leukemia: a simulation study based on phase III trial data. *Haematologica* 103, 1825–1834. [PubMed: 29954936]
- Frei E 3rd. (1985). Curative cancer chemotherapy. *Cancer Res.* 45, 6523–6537. [PubMed: 2998603]
- Gerrish PJ, and Lenski RE (1998). The fate of competing beneficial mutations in an asexual population. *Genetica* 102–103, 127–144.
- Goldberg DE, Siliciano RF, and Jacobs WR Jr. (2012). Outwitting evolution: fighting drug-resistant TB, malaria, and HIV. *Cell* 148, 1271–1283. [PubMed: 22424234]
- Goldie JH, and Coldman AJ (1979). A mathematic model for relating the drug sensitivity of tumors to their spontaneous mutation rate. *Cancer Treat. Rep* 63, 1727–1733. [PubMed: 526911]
- Goldie JH, and Coldman AJ (1984). The genetic origin of drug resistance in neoplasms: implications for systemic therapy. *Cancer Res.* 44, 3643–3653. [PubMed: 6744284]
- Gozgit JM, Chen TH, Song Y, Wardwell S, Wang F, Cai J, Li H, Edgren H, Rivera VM, and Pritchard J (2018). RET fusions observed in lung and colorectal cancers are sensitive to ponatinib. *Oncotarget* 9, 29654–29664. [PubMed: 30038711]
- Hall BG (2004). Predicting the evolution of antibiotic resistance genes. *Nat. Rev. Microbiol* 2, 430–435. [PubMed: 15100696]
- Hastings IM (2004). The origins of antimalarial drug resistance. *Trends Parasitol.* 20, 512–518. [PubMed: 15471702]

- Hauser K, Negron C, Albanese SK, Ray S, Steinbrecher T, Abel R, Chodera JD, and Wang L (2018). Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. *Commun. Biol* 1, 70. [PubMed: 30159405]
- Hochhaus A, Saglio G, Larson RA, Kim DW, Etienne G, Rosti G, De Souza C, Kurokawa M, Kalaycio ME, Hoenekopp A, et al. (2013). Nilotinib is associated with a reduced incidence of BCR-ABL mutations vs imatinib in patients with newly diagnosed chronic myeloid leukemia in chronic phase. *Blood* 121, 3703–3708. [PubMed: 23502220]
- Hughes TP, Saglio G, Quintás-Cardama A, Mauro MJ, Kim DW, Lipton JH, Bradley-Garelik MB, Ukrepec J, and Hochhaus A (2015). BCR-ABL1 mutation development during first-line treatment with dasatinib or imatinib for chronic myeloid leukemia in chronic phase. *Leukemia* 29, 1832–1838. [PubMed: 26118315]
- Iwasa Y, Nowak MA, and Michor F (2006). Evolution of Resistance During Clonal Expansion. *Genetics* 172, 2557–2566. [PubMed: 16636113]
- Jabbour E, Kantarjian H, and Cortes J (2015). Use of second- and third-generation tyrosine kinase inhibitors in the treatment of chronic myeloid leukemia: an evolving treatment paradigm. *Clin. Lymphoma Myeloma Leuk* 15, 323–334. [PubMed: 25971713]
- Kim SH, Kim D, Kim DW, Goh HG, Jang SE, Lee J, Kim WS, Kweon IY, and Park SH (2009). Analysis of Bcr-Abl kinase domain mutations in Korean chronic myeloid leukaemia patients: poor clinical outcome of P-loop and T315I mutation is disease phase dependent. *Hematol. Oncol* 27, 190–197. [PubMed: 19274615]
- Komarova N (2006). Stochastic modeling of drug resistance in cancer. *J. Theor. Biol* 239, 351–366. [PubMed: 16194548]
- Komarova NL, and Wodarz D (2005). Drug resistance in cancer: principles of emergence and prevention. *Proc. Natl. Acad. Sci. USA* 102, 9714–9719. [PubMed: 15980154]
- Lässig M, Mustonen V, and Walczak AM (2017). Predicting evolution. *Nat. Ecol. Evol* 1, 77. [PubMed: 28812721]
- LeClerc JE, Li B, Payne WL, and Cebula TA (1996). High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 274, 1208–1211. [PubMed: 8895473]
- Lind PA, Libby E, Herzog J, and Rainey PB (2019). Predicting mutational routes to new adaptive phenotypes. *Elife* 8, e38822. [PubMed: 30616716]
- Lobkovsky AE, and Koonin EV (2012). Replaying the tape of life: quantification of the predictability of evolution. *Front. Genet* 3, 246. [PubMed: 23226153]
- Luka išinová M, and Bollenbach T (2017). Toward a quantitative understanding of antibiotic resistance evolution. *Curr. Opin. Biotechnol* 46, 90–97. [PubMed: 28292709]
- Łuksza M, and Lässig M (2014). A predictive fitness model for influenza. *Nature* 507, 57–61. [PubMed: 24572367]
- Luria SE, and Delbrück M (1943). Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* 28, 491–511. [PubMed: 17247100]
- Menichincheri M, Ardini E, Magnaghi P, Avanzi N, Banfi P, Bossi R, Buffa L, Canevari G, Ceriani L, Colombo M, et al. (2016). Discovery of Entrectinib: A New 3-Aminoindazole As a Potent Anaplastic Lymphoma Kinase (ALK), c-ros Oncogene 1 Kinase (ROS1), and Pan-Tropomyosin Receptor Kinases (Pan-TRKs) inhibitor. *J. Med. Chem* 59, 3392–3408. [PubMed: 27003761]
- Michor F, Hughes TP, Iwasa Y, Branford S, Shah NP, Sawyers CL, and Nowak MA (2005). Dynamics of chronic myeloid leukaemia. *Nature* 435, 1267–1270. [PubMed: 15988530]
- Mok TS, Wu YL, Ahn MJ, Garassino MC, Kim HR, Ramalingam SS, Shepherd FA, He Y, Akamatsu H, Theelen WSME, et al.; AURA3 Investigators (2017). Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer. *N. Engl. J. Med* 376, 629–640. [PubMed: 27959700]
- O'Hare T, Shakespeare WC, Zhu X, Eide CA, Rivera VM, Wang F, Adrian LT, Zhou T, Huang WS, Xu Q, et al. (2009). AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. *Cancer Cell* 16, 401–412. [PubMed: 19878872]
- Payne JL, Menardo F, Trauner A, Borrell S, Gygli SM, Loiseau C, Gagneux S, and Hall AR (2019). Transition bias influences the evolution of antibiotic resistance in *Mycobacterium tuberculosis*. *PLoS Biol.* 17, e3000265. [PubMed: 31083647]

- Peng B, Hayes M, Resta D, Racine-Poon A, Druker BJ, Talpaz M, Sawyers CL, Rosamilia M, Ford J, Lloyd P, and Capdeville R (2004). Pharmacokinetics and pharmacodynamics of imatinib in a phase I trial with chronic myeloid leukemia patients. *J. Clin. Oncol* 22, 935–942. [PubMed: 14990650]
- Peng B, Lloyd P, and Schran H (2005). Clinical pharmacokinetics of imatinib. *Clin. Pharmacokinet* 44, 879–894. [PubMed: 16122278]
- Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim DW, Ou SI, Pérol M, Dziadziuszko R, Rosell R, et al. (2017). Alectinib versus Crizotinib in Untreated ALK-Positive Non-Small-Cell Lung Cancer. *N. Engl. J. Med* 377, 829–838. [PubMed: 28586279]
- Pritchard JR, Lauffenburger DA, and Hemann MT (2012). Understanding resistance to combination chemotherapy. *Drug Resist. Updat* 15, 249–257. [PubMed: 23164555]
- Pritchard JR, Bruno PM, Gilbert LA, Capron KL, Lauffenburger DA, and Hemann MT (2013). Defining principles of combination drug mechanisms of action. *Proc. Natl. Acad. Sci. USA* 110, E170–E179. [PubMed: 23251029]
- Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, and Shakhnovich EI (2016). Biophysical principles predict fitness landscapes of drug resistance. *Proc. Natl. Acad. Sci. USA* 113, E1470–E1478. [PubMed: 26929328]
- Shah NP, Tran C, Lee FY, Chen P, Norris D, and Sawyers CL (2004). Overriding Imatinib Resistance with a Novel ABL Kinase Inhibitor. *Science* 305, 399–401. [PubMed: 15256671]
- Stein AM, Bottino D, Modur V, Branford S, Kaeda J, Goldman JM, Hughes TP, Radich JP, and Hochhaus A (2011). BCR-ABL transcript dynamics support the hypothesis that leukemic stem cells are reduced during imatinib treatment. *Clin. Cancer Res* 17, 6812–6821. [PubMed: 21903771]
- Stoltzfus A, and McCandlish DM (2017). Mutational Biases Influence Parallel Adaptation. *Mol. Biol. Evol* 34, 2163–2172. [PubMed: 28645195]
- Subbiah V, Gainor JF, Rahal R, Brubaker JD, Kim JL, Maynard M, Hu W, Cao Q, Sheets MP, Wilson D, et al. (2018). Precision Targeted Therapy With BLU-667 for RET-Driven Cancers. *Cancer Discov.* 8, 836–849. [PubMed: 29657135]
- Wylie AA, Schoepfer J, Jahnke W, Cowan-Jacob SW, Loo A, Furet P, Marzinzik AL, Pelle X, Donovan J, Zhu W, et al. (2017). The allosteric inhibitor ABL001 enables dual targeting of BCR-ABL1. *Nature* 543, 733–737. [PubMed: 28329763]
- Zhang S, Anjum R, Squillace R, Nadworny S, Zhou T, Keats J, Ning Y, Wardwell SD, Miller D, Song Y, et al. (2016). The Potent ALK Inhibitor Brigatinib (AP26113) Overcomes Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in Preclinical Models. *Clin. Cancer Res* 22, 5527–5538. [PubMed: 27780853]

**Highlights**

- Multi-scale, mechanistic models can predict drug resistance epidemiology
- For low-heterogeneity diseases, substitution likelihood is a significant predictor
- Resistance to adjuvant therapy drugs may be shaped by substitution likelihood
- Drugs that prioritize must-hit variants can reduce the incidence of resistance



**Figure 1. A Salt Bridge in ABL1 Suggests that Clinical Abundance May Not Be Predicted by the Amount of Drug Resistance Conferred**

(A) Schematic of factors affecting the evolutionary dynamics of drug resistance. Without transmission, intra-tumoral variables are the only factors involved in cancer evolutionary dynamics and are limited to the host level.

(B) ABL1 crystal structure. Ribbon diagram of secondary structure is shown. Image is zoomed in on the kinase P loop. Loss of the E255-K247 salt bridge is associated with imatinib resistance.

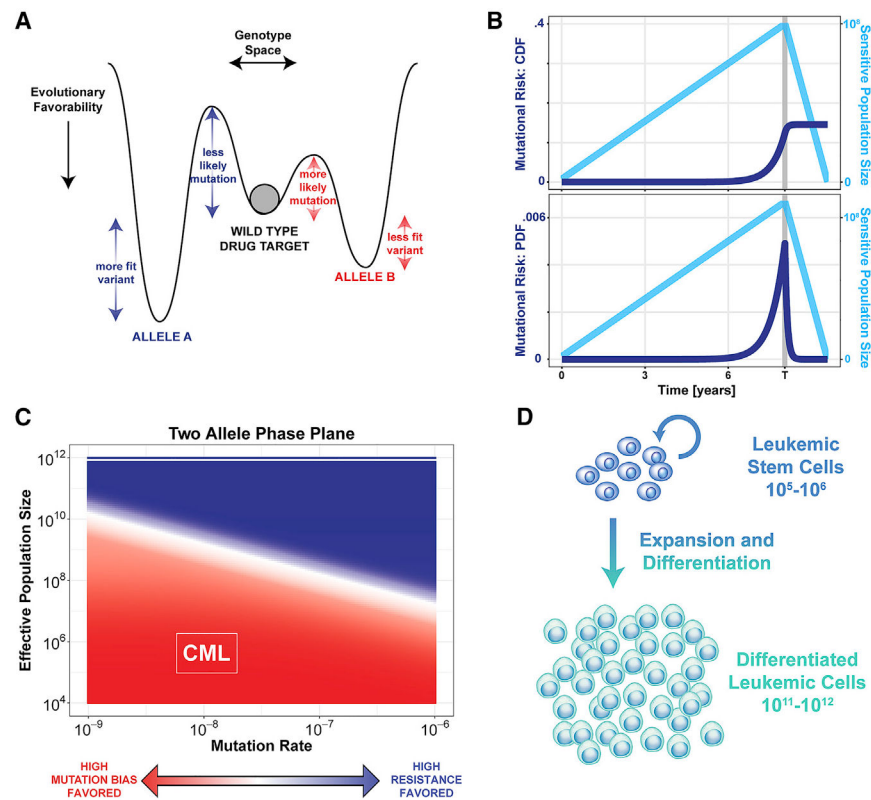
(C) Prevalence of E255K/E255V mutations in six imatinib clinical trials. A cross-trial sum is included; the p value is for the chi-square test.

(D) Imatinib  $IC_{50}$  curves for BCR-ABL-transformed BaF3 cells. Relative viability is measured by Cell-Titer Glo relative to a DMSO control.  $n = 3$  biological replicates per concentration; error bars are SDs.

(E) Relative growth rates of BCR-ABL BaF3 variants. Each dot is an independent transduction and selection.  $n = 11-14$  biological replicates; error bars are SDs.

(F) Top: codon structure of E255 and various measurements of nucleotide substitution frequencies. These measurements include counts of all rare substitutions across the ABL1 gene identified in the Broad ExAC database, rare noncoding substitutions in ABL1 identified in the Broad ExAC database, and substitutions across the exome in CML patients. Lower: different measurements of substitution likelihood were highly correlated.





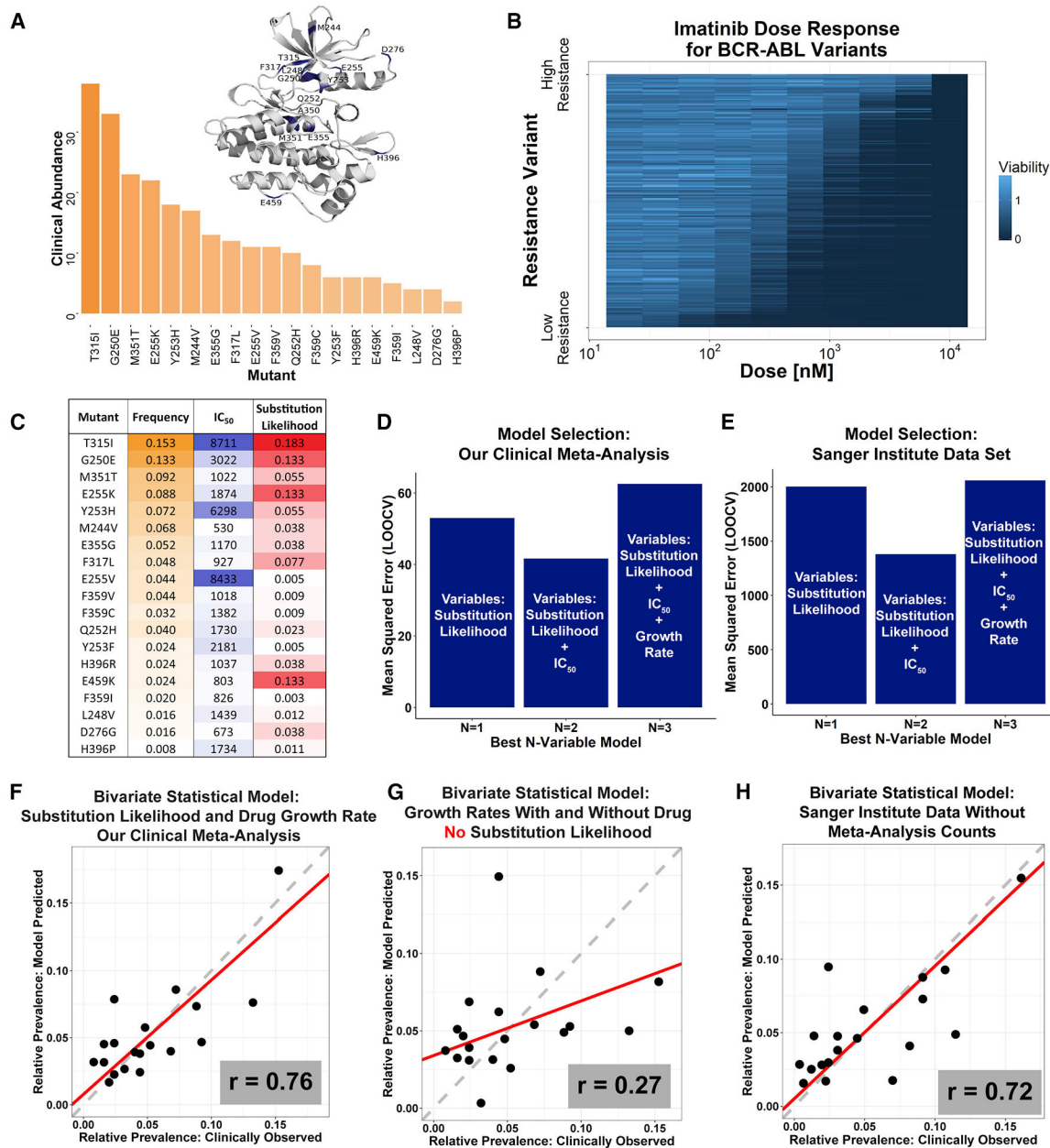
**Figure 2. An Analytical Model of Stochastic Dynamics Identifies in which Resistance Evolution Favors More Likely Alleles**

(A) Evolutionary landscape for a theoretical drug target gene with two potential resistance alleles. Allele A is assigned high fitness and low probability; allele B is assigned low fitness and high probability.

(B) Cumulative distribution function (CDF) (top plot) indicates the probability that a mutation has occurred by time  $t$ . As  $t$  approaches infinity, the CDF approaches the overall probability of mutation used in the two-way table. The probability density function (PDF) (bottom plot) indicates the instantaneous risk of mutation at time  $t$ . This is used to determine the dominant allele in competition, when both alleles are present. The time  $T$  indicates the time of detection and treatment initiation. See STAR Methods for more details.

(C) Parameter space indicating the results of our probability model across many mutation rates and effective population sizes. Color indicates whether allele A (dark blue) or allele B (red) is more likely to drive relapse. In regions where allele B is more dominant, we expect mutation bias to be a primary evolutionary force.

(D) Schematic of general leukemic cell population hierarchy. Only mutations in leukemic stem cells can form stable resistance clones, effectively limiting the population size to  $10^5$ – $10^6$ .



**Figure 3. Epidemiological Incidences of ABL1 Resistance Mutations Are Best Predicted by How Likely They Are**

(A) Crystal structure ribbon diagram of the ABL1 kinase domain and distribution of the 19 most prevalent BCR-ABL resistance mutants. These 19 variants account for approximately 95% of resistance mutations observed clinically in the six studies from Figure 1C.

(B) Drug-dose response was measured by Cell-Titer Glo and is plotted in the heatmap.  $n = 3$  independent infections for all 20 cell lines (wild type [WT] and mutants). Each BCR-ABL BaF3 line was dosed with 11 serial dilutions of imatinib in triplicate. All cell lines are ordered by sensitivity. For each line, three biological replicates were conducted, each with two technical replicates.

(C) Table that includes values used to build the final regression models: the frequency of each resistant mutant as determined by our clinical meta-analysis, imatinib  $IC_{50}$  values (in nanomolar) normalized for genetic background, and substitution likelihood calculated from analysis of the Broad ExAC data in Figure 1F.

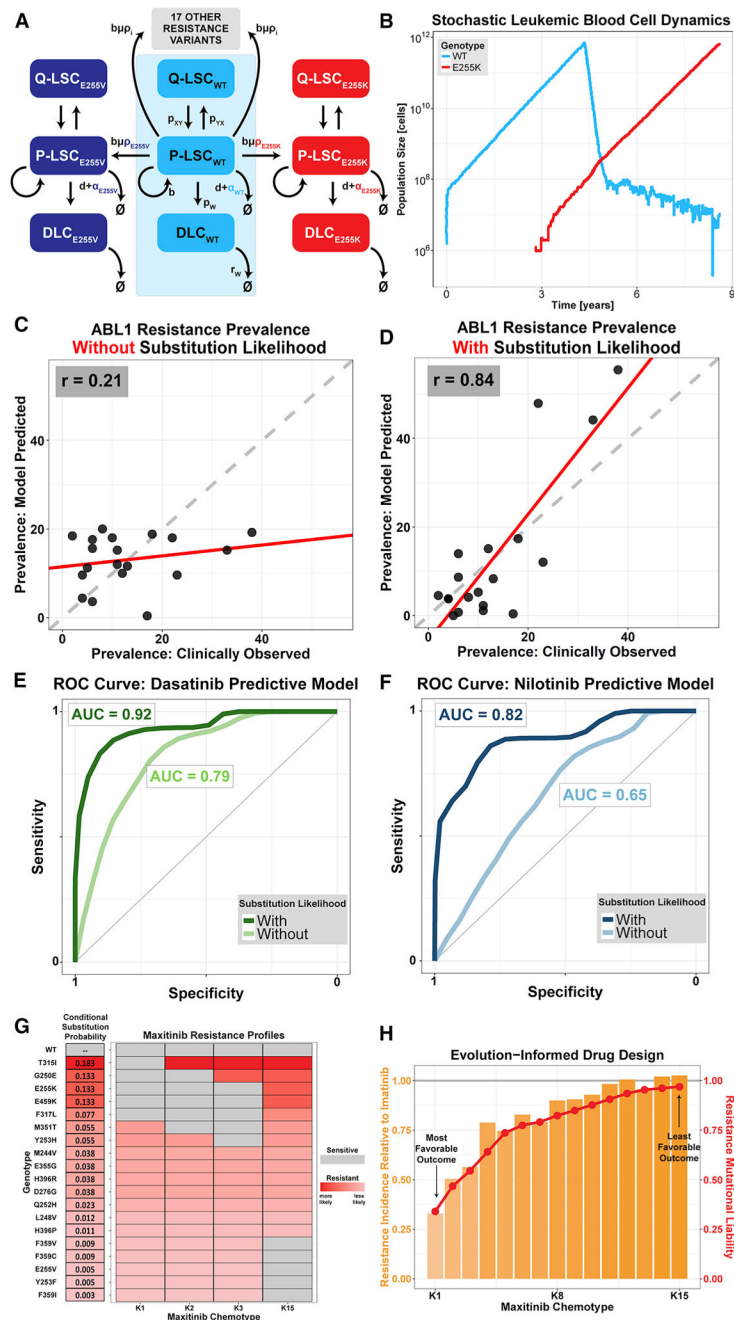
(D) Model selection for negative binomial regression. LOOCV was used to estimate the test error of the N-variable model with the lowest Akaike information criterion (AIC) for each N. The N = 1 model includes substitution bias, the N = 2 model includes substitution bias and  $IC_{50}$ , and the N = 3 model includes substitution bias,  $IC_{50}$ , and growth rate in the absence of drug. The N = 2 model had the lowest estimated mean square error. See .rmd on GitHub for more details.

(E) Model selection for regression (as in D) using the Sanger dataset.

(F) Observed versus predicted plot for a regression model of clinical mutation prevalence (determined by our clinical meta-analysis) regressed against growth rate in the presence of drug and substitution likelihood (which is the final model). Points are specific ABL1 mutations; x values are their observed frequency, and y values are their frequency predicted by the model. The Pearson correlation for observed and predicted prevalence is 0.76.

(G) Observed versus predicted plot (as in F) for the regression model of clinical mutation prevalence built only on the growth rates in the presence and absence of imatinib. The Pearson correlation is 0.27.

(H) Observed versus predicted plot (as in F) using the Sanger data with the counts from our meta-analysis omitted, to consider the possibility that the Sanger data included all counts from our curated dataset. The Pearson correlation is 0.72.



**Figure 4. A Stochastic, First-Principle, Multi-mutation Model of Imatinib Treatment Predicts the Clinical Prevalence of Resistance Mutations across ABL1**

(A) Schematic of a stochastic CML evolutionary dynamic model. The initial deterministic model of three differential equations (shaded in light blue) is from Fassoni et al. (2018) and is fit to phase III clinical data. We reformulated a stochastic version of 60 differential equations parameterized from Fassoni et al. (2018) and our clinical data. Leukemic stem cells alternate between the proliferating state (P-LSC) and the quiescent state (Q-LSC). P-LSCs give rise to differentiated leukemic cells (DLCs). P-LSCs may also spawn a resistant subclone P-LSC<sub>i</sub> when dividing. The allele-specific mutational probability is given by  $p_i$ .

We added the ability for all resistance mutations to occur, such that there are 20 sets of differential equations with three populations per mutant. The system is solved stochastically.

(B) Example stochastic simulation of the model described in (A).

(C) Simulation results for the stochastic model without mutation bias (uniform  $\rho_i$ ). The Pearson correlation between observed and predicted prevalence is  $r = 0.21$ .

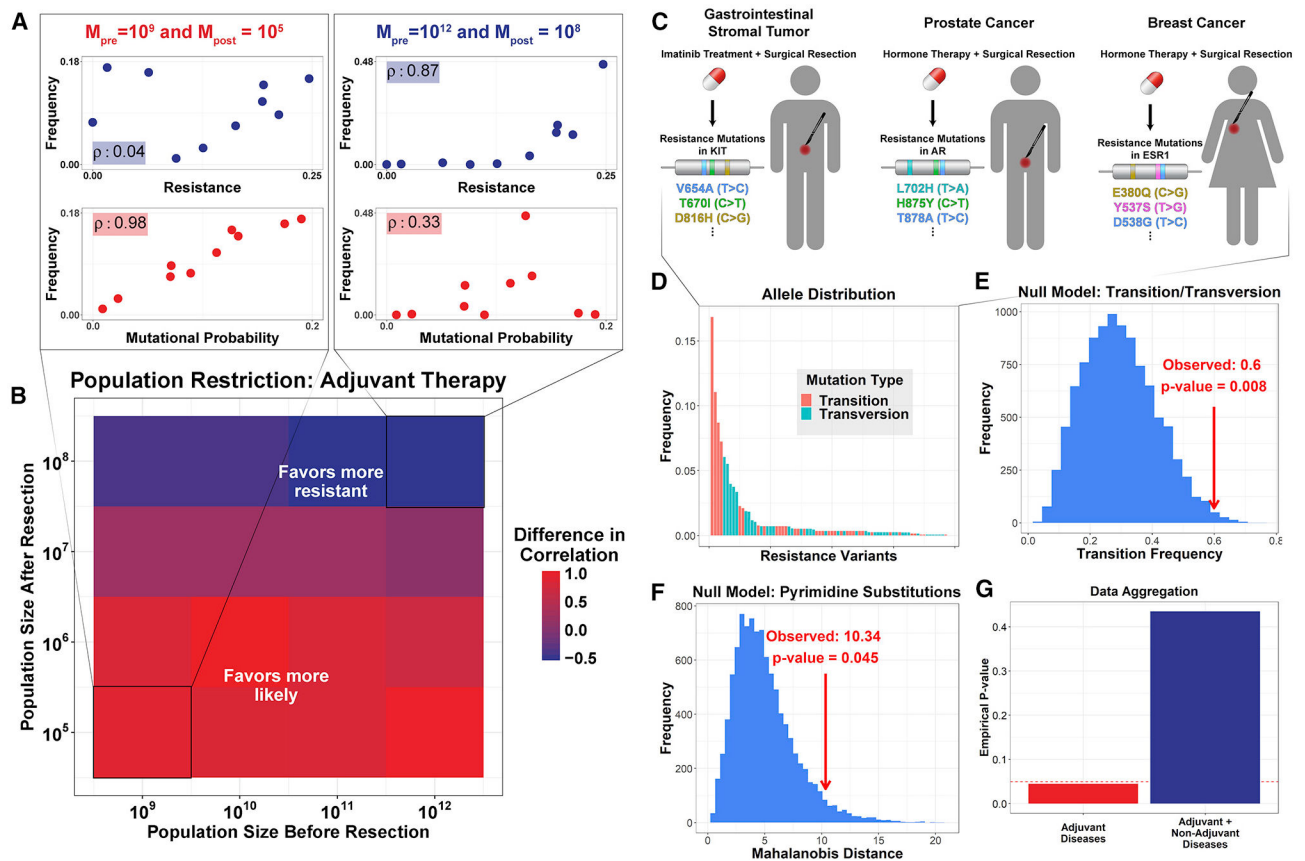
(D) Simulation results for the stochastic model with mutation bias (allele-specific  $\rho_i$ ). The Pearson correlation is  $r = 0.84$ .

(E) ROC curves indicating the accuracy of simulation results for the stochastic model using parameters for dasatinib. The model-predicted allele frequency was used to train a binary classifier of clinical resistance, and its predictive power was evaluated against the true positive and negative classes as identified in a frontline dasatinib clinical trial. The stochastic model that included mutation bias (AUC = 0.92) outperformed that without bias (AUC = 0.79).

(F) ROC curve for simulations using parameters for nilotinib, as in (E). Again, the model with mutation bias was more predictive than the one without mutation bias (AUC = 0.82 with bias versus AUC = 0.65 without bias).

(G) Resistance profiles for versions of a hypothetical drug, maxitinib. Maxitinib K1 targets the first through fifth most likely mutants, K2 targets the second through sixth most likely mutants, and so on.

(H) Maxitinib simulation results. Orange bars represent the resistance incidence of each maxitinib chemotype relative to imatinib. Red points indicate mutational liability, defined as the sum of conditional substitution likelihoods of mutations that confer resistance to each chemotype.



**Figure 5. Mutation Bias Affects the Epidemiology of Resistance for Diseases Treated in the Adjuvant Setting**

(A) Adjuvant therapy evolutionary model example results. Simulation results for population before resection  $M_{pre} = 10^9$  and after resection  $M_{post} = 10^5$  (left) and for  $M_{pre} = 10^{12}$  and  $M_{post} = 10^8$  (right). Points represent specific resistance variants, and  $p$  values are the Spearman rank correlation between frequency and degree of resistance (top) and frequency and substitution likelihood (bottom).

(B) Summary of simulation results for various values of  $M_{pre}$  and  $M_{post}$ . Colors represent the difference in Spearman correlation (substitution likelihood  $\rho$  — resistance  $\rho$ ) for each set of parameters.

(C) Schematic detailing clinical meta-analysis. For drug targets in GISTs, prostate cancer, and breast cancer, acquired resistance mutations were tallied and classified by pyrimidine substitution.

(D) Observed resistance allele distribution for the three cancer types. Colors indicate whether the nucleotide substitution associated with each mutation is a transition or a transversion.

(E) Distribution simulated by reassigning the transition-transversion class of each variant. The transition frequency in each simulation was calculated. The histogram shows the distribution of those frequencies. The red arrow indicates the transition frequency for the observed data (0.6). The associated empirical  $p$  value is 0.008, suggesting that the observed data cannot be explained by a null model with no differences in probability between transitions and transversions.

(F) Distribution as in (E), but with mutation classes (e.g. C > A, C > G, etc.).

(G) Combining the counts from the three adjuvant diseases with those of ALK+ and EGFR+ NSCLC (two nonadjuvant indications) decreases the significance detected by the model.

This suggests that the observed significance in the adjuvant diseases is not an artifact of aggregating data and large sample size.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
CellTiter Glo	Promega	G7570
Deposited Data		
BCR-ABL BaF3 IC <sub>50</sub> - data for each replicate	This Paper	<a href="https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/blob/master/Figures/Figure3/Heatmaps/IC50HeatMap.csv">https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/blob/master/Figures/Figure3/Heatmaps/IC50HeatMap.csv</a>
BCR-ABL BaF3 drug free growth - data for each replicate	This Paper	<a href="https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/blob/master/Figures/Figure3/Heatmaps/GrowthCurveHeatMap.csv">https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/blob/master/Figures/Figure3/Heatmaps/GrowthCurveHeatMap.csv</a>
Experimental Models: Cell Lines		
BaF3 Cells	DSMZ	ACC 300
Recombinant DNA		
pLVX-IRES-Puro	Clontech	Catalog number: 632183
BCR-ABL Gene Synthesis	Genscript	Accession number: EU216071.1
Software and Algorithms		
Code for CML stochastic model	This Paper	<a href="https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure4/Figure4CDEF">https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure4/Figure4CDEF</a>
Code for adjuvant therapy stochastic model	This paper	<a href="https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure5/Figure6AB">https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure5/Figure6AB</a>
Code for mutation bias detection from allele distribution	This paper	<a href="https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure5/FigureDE">https://github.com/pritchardlabatpsu/PredictiveResistanceEvolution/tree/master/Figures/Figure5/FigureDE</a>