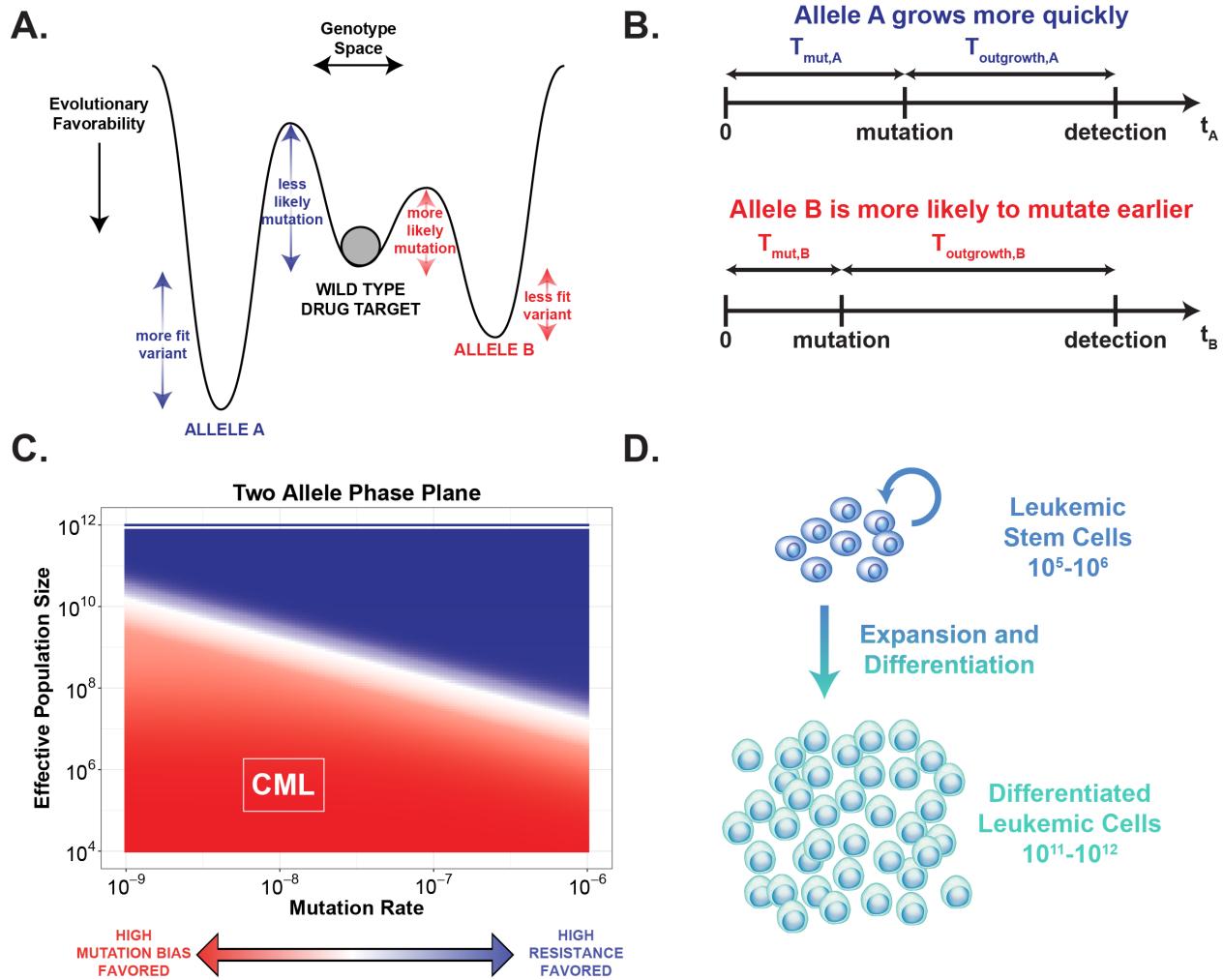


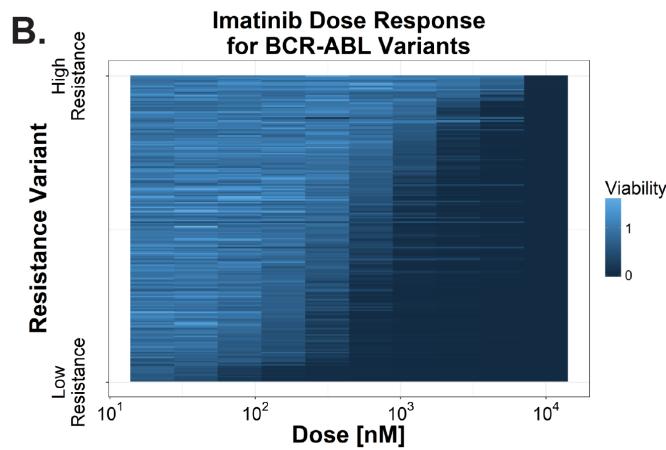
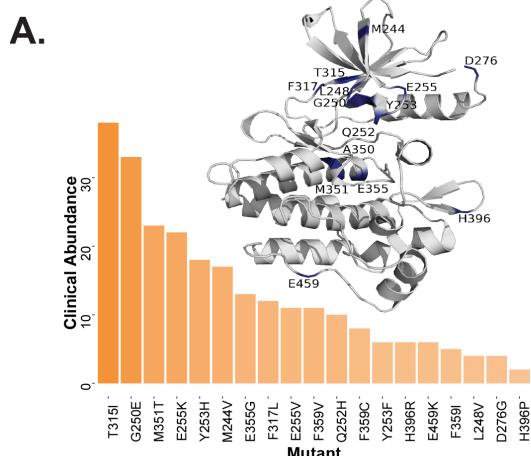
**Figure 1: A salt bridge in ABL1 suggests that clinical abundance may not be predicted by the amount of drug resistance conferred.**

**(A)** 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 E225-K247 salt bridge is associated with imatinib resistance. **(C)** Prevalence of E255K/V mutations in six imatinib clinical trials. A cross-trial sum is included, p-value is for chi-square test. **(D)** Imatinib IC<sub>50</sub> curves for BCR-ABL transformed BaF3 cells. Relative viability is measured by Cell-Titer Glo relative to a DMSO control. N=3 per concentration, error bars are standard deviations. **(E)** Relative growth rates of BCR-ABL BaF3 variants. Each dot is an independent transduction and selection. N=11-14, error bars are standard deviations. **(F)** *Upper:* Codon structure of E255 and various measurements of nucleotide substitution frequencies. These measurements include counts of substitutions across the ABL1 gene identified in the Broad ExAC database; 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 where survival of the likeliest can occur.**

(A) Evolutionary landscape for a theoretical drug target gene with two potential resistance alleles. Allele A is assigned a high fitness and low probability; Allele B is assigned a low fitness and high probability. (B) A schematic of a general timeline for mutation and outgrowth for either allele given the assigned evolutionary profiles. In cases where both mutations occur, the first resistant clone to reach detection drives relapse. (C) A phase plane of 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  $10^5$ - $10^6$ .

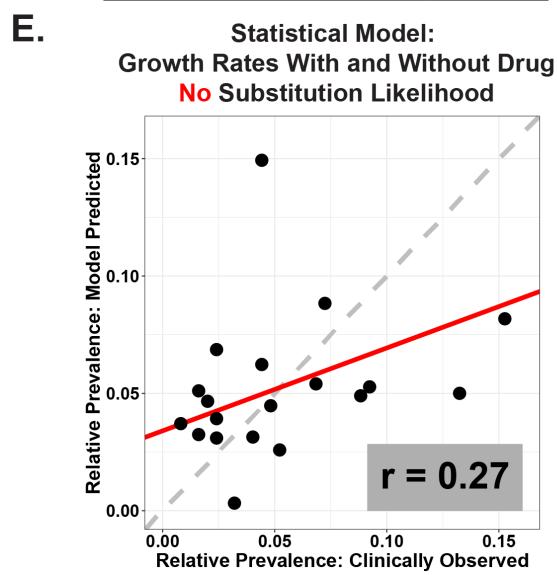
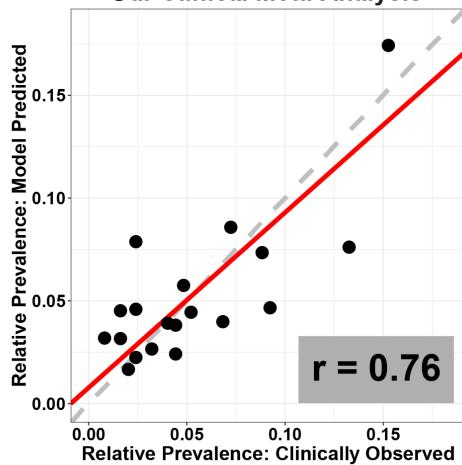


**C.**

Mutant	Frequency	IC <sub>50</sub>	Mutation Probability
T315I	0.153	8711	0.183
G250E	0.133	3022	0.133
M351T	0.092	1022	0.055
E255K	0.088	1874	0.133
Y253H	0.072	6298	0.055
M244V	0.068	530	0.038
E355G	0.052	1170	0.038
F317L	0.048	927	0.077
E255V	0.044	8433	0.005
F359V	0.044	1018	0.009
F359C	0.032	1382	0.009
Q252H	0.040	1730	0.023
Y253F	0.024	2181	0.005
H396R	0.024	1037	0.038
E459K	0.024	803	0.133
F359I	0.020	826	0.003
L248V	0.016	1439	0.012
D276G	0.016	673	0.038
H396P	0.008	1734	0.011

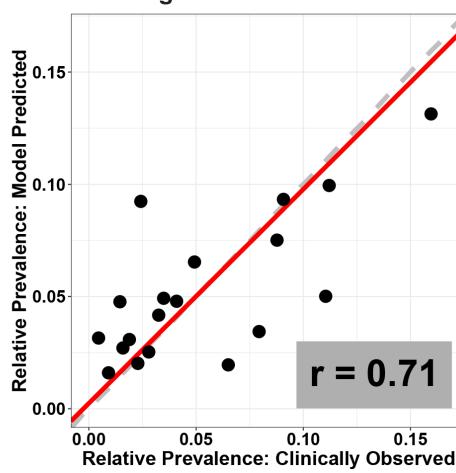
**D.**

**Statistical Model:  
Substitution Likelihood and Drug Growth Rate  
Our Clinical Meta-Analysis**



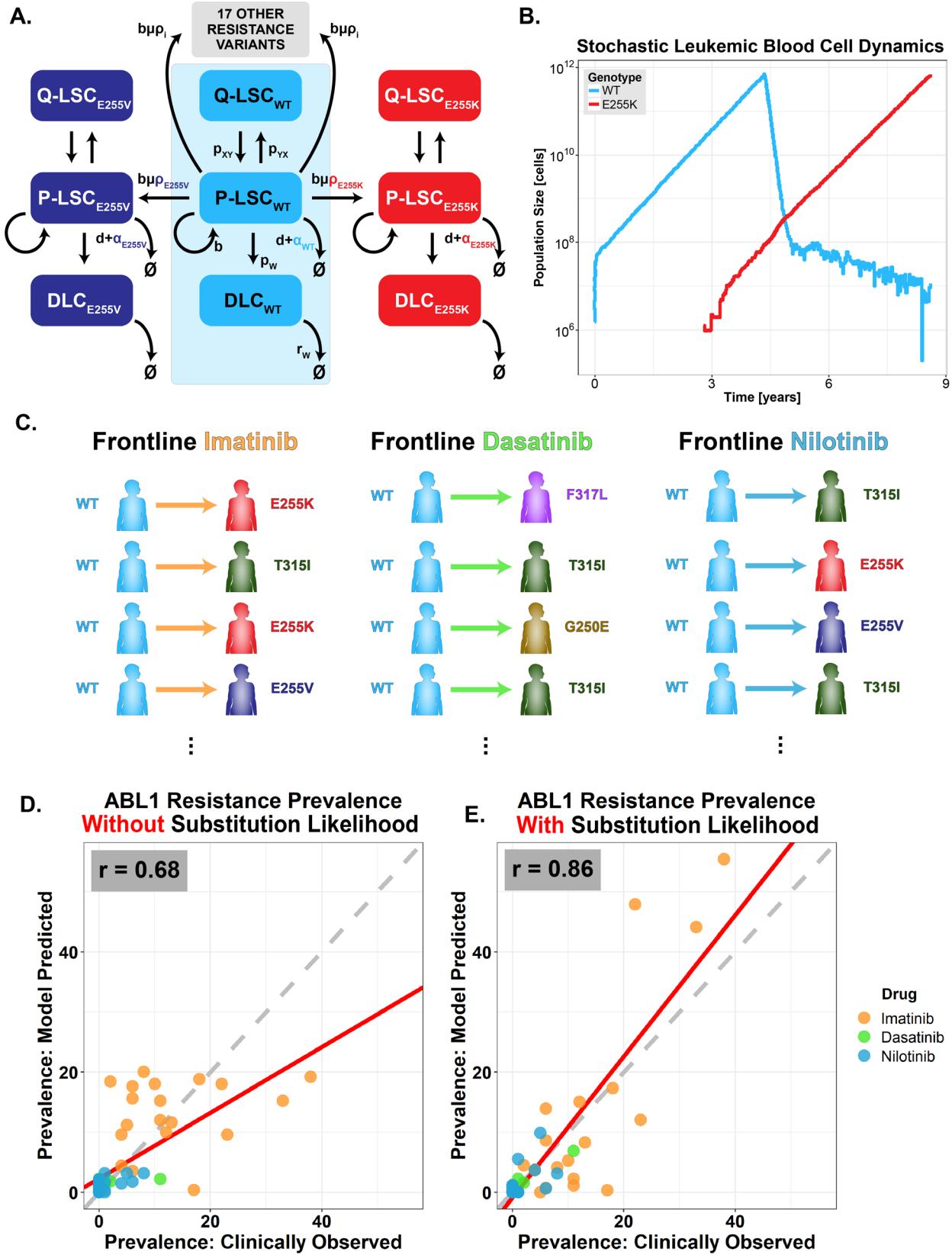
**F.**

**Statistical Model:  
Substitution Likelihood and Drug Growth Rate  
Sanger Institute Data Set**



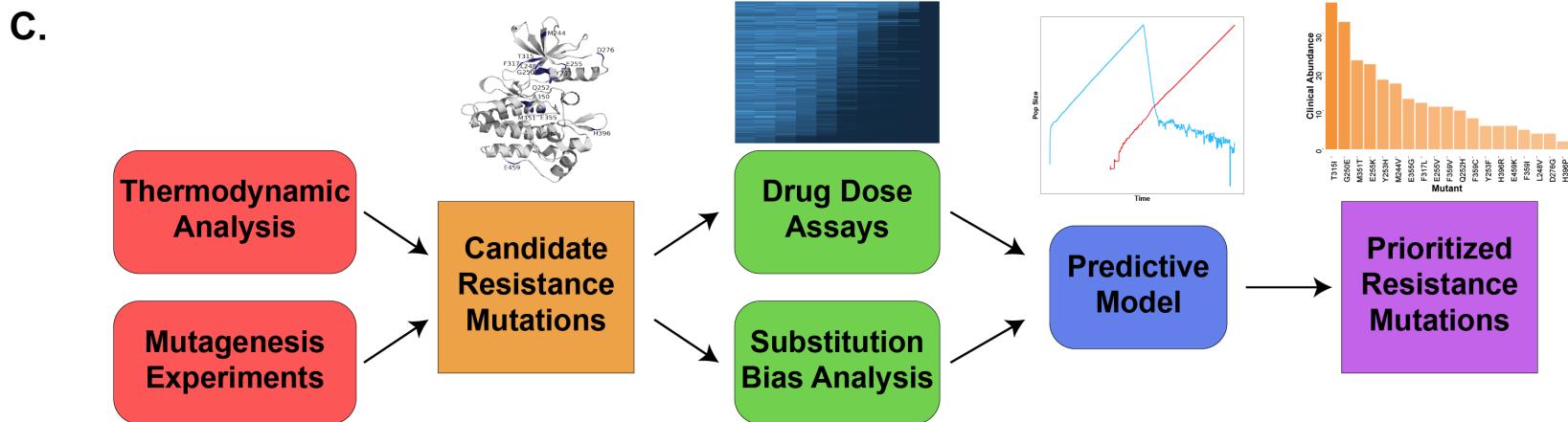
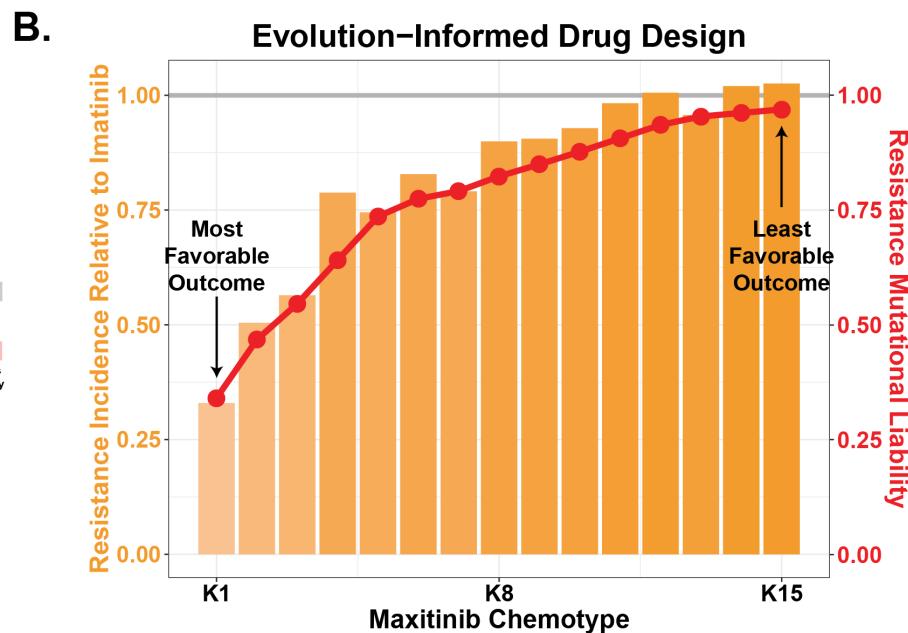
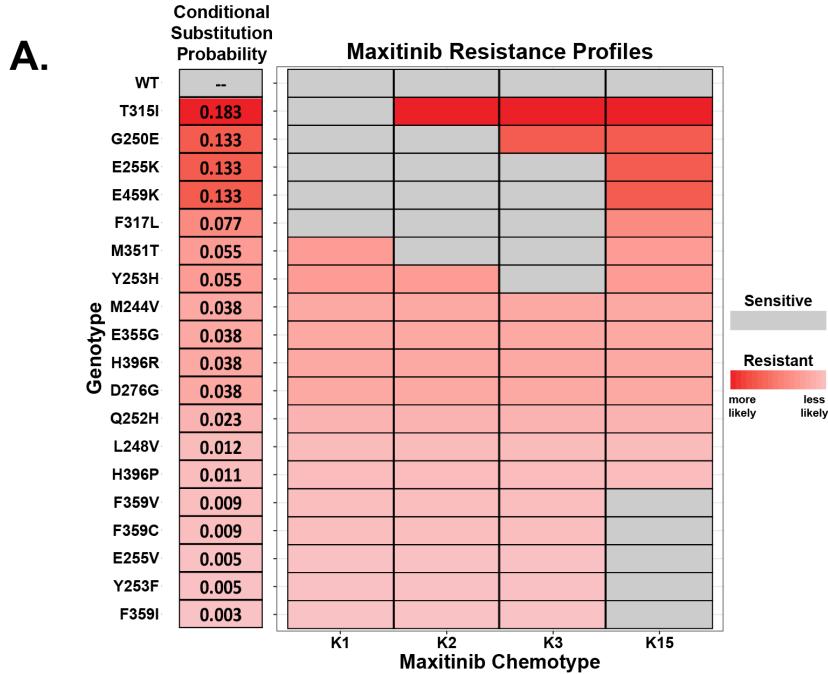
**Figure 3: Epidemiologic incidences of ABL1 resistance mutations are best predicted by how likely they are.**

**(A)** 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 (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. Raw data and code are available on GitHub. **(C)** A table that includes values used to build regression models: the frequency of each resistant mutant as determined by our clinical meta-analysis, imatinib IC<sub>50</sub>s (in nM) normalized for genetic background, and substitution likelihood calculated from analysis of the Broad ExAC data in Figure 1F. **(D)** 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. Pearson correlation for observed and predicted prevalences is  $r = 0.76$ . **(E)** Observed versus predicted plot (as in Figure 3D) for the regression model of clinical mutation prevalence built only on the growth rates in the presence and absence of imatinib. Pearson correlation  $r = 0.27$ . **(F)** Observed versus predicted plot (as in Figure 3D) for independent Sanger Institute data set. Pearson correlation  $r = 0.71$ .



**Figure 4: A stochastic, first principles, multi-mutation model of imatinib treatment predicts the clinical prevalence of resistance mutations across ABL1.**

**(A)** Schematic of stochastic CML evolutionary dynamic model. The initial deterministic model of three differential equations (shaded in blue) is from Fassoni et al. 2018<sup>38</sup> and is fit to phase 3 clinical data. We reformulated a stochastic version of 60 differential equations parameterized from Fassoni et al. and our clinical data. Leukemic stem cells alternate between proliferating (P-LSC) and quiescent state (Q-LSC). P-LSC give rise to differentiated leukemic cells (DLC). P-LSCs may also spawn a resistant subclone P-LSC<sub>i</sub> when dividing. The allele-specific mutational probability is given by  $\rho_i$ . Note that we added the ability for all 19 resistance mutations to occur, such that there are 20 sets of differential equations with three populations per mutant. The system is solved stochastically. **(B)** An example stochastic simulation of the model described in Figure 4A. **(C)** Simulations were conducted (as in Figure 4B) 10,000 times for each of three BCR-ABL inhibitors: imatinib, dasatinib, and nilotinib. Resistance alleles were tallied across simulations for the three drugs. **(D)** Simulation results for the stochastic model without mutation bias (uniform  $\rho_i$ ). The Pearson correlation between observed and predicted prevalences is  $r = 0.68$ . **(E)** Simulation results for the stochastic model with mutation bias (allele-specific  $\rho_i$ ). The Pearson correlation is  $r = 0.86$ .

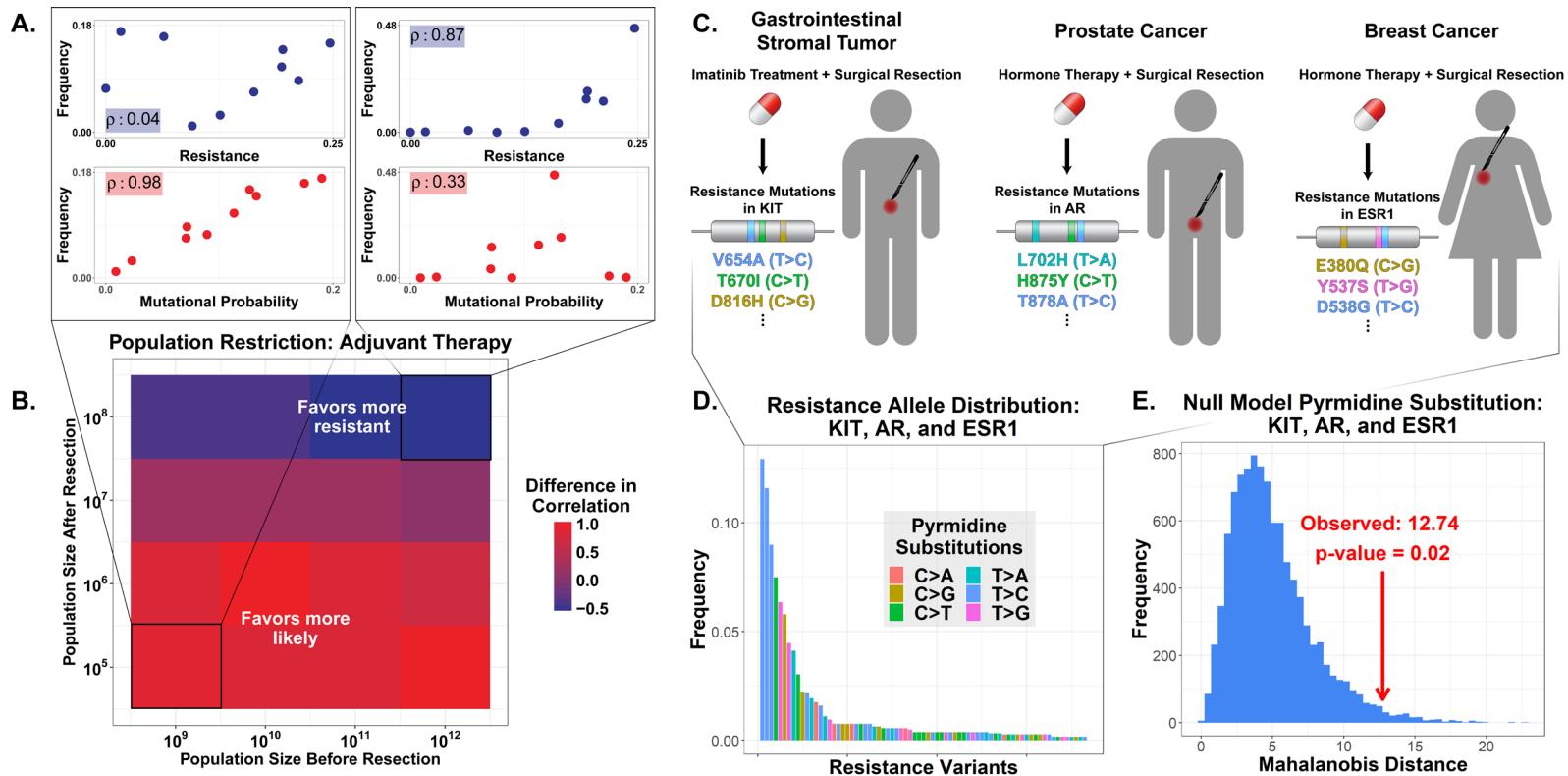


**Figure 5: An evolution-guided approach drug design is predicted to minimize resistance prevalence**

**(A)** 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; and so on.

**(B)** 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.

**(C)** A proposed workflow for evolution guided drug design. Potential resistance mutations could be generated by structure driven simulation, unbiased mutagenesis, or CRISPR base editors. Mutations can then be analyzed as single IC<sub>50</sub>s or in a pooled format. Indication specific information on the substitution biases would require the mutational signature for a given disease. Fitness and bias estimates could then be coupled into a mechanistic model of drug resistance that would predict the clinically most abundant resistance mutations.



**Figure 6: Restricted genetic heterogeneity in adjuvant therapy gives rise to survival of the likeliest**

**(A)** Adjuvant therapy evolutionary model example results. Simulation results for population before resection  $M_{\text{pre}}=10^9$  and after resection  $M_{\text{post}}=10^5$  (*left*) and for  $M_{\text{pre}}=10^{12}$  and  $M_{\text{post}}=10^8$  (*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  $M_{\text{pre}}$  and  $M_{\text{post}}$ . Colors represent the difference in correlation (substitution likelihood p – resistance p) for each set of parameters. **(C)** Schematic detailing clinical meta-analysis. For drug targets in GIST, 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 the pyrimidine substitution associated with each mutation. **(E)** The distribution in Figure 6D was simulated by reassigning the mutation class of each variant. Each simulations' Mahalanobis distance from the distribution of all simulations was calculated. The histogram shows the distribution of those distances. The red arrow indicates the Mahalanobis distance of the observed data from the simulated null distribution (distance = 12.74). The empirical p-value of the real-world data is 0.02, suggesting that the observed data cannot be explained by a null model with no differences in substitution likelihood.