# CRISPR-Cas9–mediated saturated mutagenesis screen predicts clinical drug resistance with improved accuracy

Leyuan Ma[a,b], Jeffrey I. Boucher[c], Janet Paulsen[c], Sebastian Matuszewski[d,e], Christopher A. Eide[f,g], Jianhong Ou[a], Garrett Eickelberg[h], Richard D. Press[h], Lihua Julie Zhu[a,i], Brian J. Druker[f,g], Susan Branford[j,k,l,m], Scot A. Wolfe[a,c], Jeffrey D. Jensen[d,e,1], Celia A. Schiffer[c], Michael R. Green[a,b,2], and Daniel N. Bolon[c,2]

[a]Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA 01605; [b]Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, MA 01605; [c]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605; [d]School of Life Sciences, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland; [e]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; [f]Knight Cancer Institute, Division of Hematology and Medical Oncology, Oregon Health & Science University, Portland, OR 97239; [g]Howard Hughes Medical Institute, Oregon Health & Science University, Portland, OR 97239; [h]Knight Cancer Institute, Department of Pathology, Oregon Health & Science University, Portland, OR 97239; [i]Programs in Molecular Medicine and Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605; [j]Department of Genetics and Molecular Pathology, Centre for Cancer Biology, SA Pathology, Adelaide, SA 5000, Australia; [k]School of Pharmacy and Medical Science, University of South Australia, Adelaide, SA 5000, Australia; [l]School of Medicine, University of Adelaide, Adelaide, SA 5005; and [m]School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

Contributed by Michael R. Green, September 19, 2017 (sent for review May 18, 2017; reviewed by Yana Bromberg and Mikko Taipale)

**Developing tools to accurately predict the clinical prevalence of drug-resistant mutations is a key step toward generating more effective therapeutics. Here we describe a high-throughput CRISPR-Cas9–based saturated mutagenesis approach to generate comprehensive libraries of point mutations at a defined genomic location and systematically study their effect on cell growth. As proof of concept, we mutagenized a selected region within the leukemic oncogene *BCR-ABL1*. Using bulk competitions with a deep-sequencing readout, we analyzed hundreds of mutations under multiple drug conditions and found that the effects of mutations on growth in the presence or absence of drug were critical for predicting clinically relevant resistant mutations, many of which were cancer adaptive in the absence of drug pressure. Using this approach, we identified all clinically isolated BCR-ABL1 mutations and achieved a prediction score that correlated highly with their clinical prevalence. The strategy described here can be broadly applied to a variety of oncogenes to predict patient mutations and evaluate resistance susceptibility in the development of new therapeutics.**

BCR-ABL | CRISPR-Cas9–based genome editing | drug resistance | saturated mutagenesis | tyrosine kinase inhibitors

The development of drug resistance limits the effectiveness of many therapeutic strategies and has a tremendous impact on disease progression and outcomes in patients (1, 2). Drug resistance has been challenging to address, in part because it is difficult to predict the mutations that will contribute to patient health. Therefore, the development of tools and strategies to predict the clinical prevalence of drug-resistant mutations in patients holds great promise for improving our ability to generate effective therapeutics with reduced susceptibility to resistance.

High-throughput random mutagenesis screens represent a promising approach for identifying mutations that lead to drug resistance. However, they have often identified many more resistant mutations than are observed in patients (3, 4), raising the question of the clinical significance of their findings. In addition, substantial variation has been observed between independent measurements in mutagenesis screens in mammalian cells, presumably due to gene expression variation caused by random viral integration, thus requiring multiple independent measurements to reliably distinguish signal from noise (5).

To improve the consistency and efficiency of in vitro mammalian screens, we revised our previously developed saturated mutational scanning approach called exceedingly meticulous and parallel investigation of randomized individual codons (EMPIRIC) (6, 7), a method that systematically generates and studies the impact of individual amino acid changes, which are often highly relevant to mammalian genetic disorders, including cancer. We sought to improve EMPIRIC by: (*i*) reducing variation by targeting mutations to a defined genomic location using CRISPR-Cas9–based genome editing, and (*ii*) facilitating multiple independent measurements for each amino acid change by utilizing a barcoding strategy; the use of multiple barcodes for each amino acid variant provides internal independent measurements within each experimental replicate. Here we describe this improved method, which we refer to as barcoded introns in the genome (BIG) EMPIRIC, and validate its application by mutagenizing a selected region within the leukemic oncogene *BCR-ABL1*.

## Significance

Many therapeutic strategies are hampered by the development of drug resistance. High-throughput random mutagenesis screens represent a promising approach for identifying mutations that lead to drug resistance, but have often identified more resistant mutations than are observed in patients, raising questions of their clinical significance. We developed an improved high-throughput mutagenesis screening approach that uses CRISPR-Cas9–based genome editing to generate comprehensive libraries of point mutations at a defined genomic location and systematically study their effect on cell growth. As proof-of-concept, we show our approach accurately predicts the clinical prevalence of drug-resistant mutations in the oncogene *BCR-ABL*. Our approach can be broadly applied to a variety of oncogenes and represents a new strategy for evaluating resistance susceptibility during drug development.

Author contributions: L.M., J.I.B., S.A.W., M.R.G., and D.N.B. designed research; L.M., J.I.B., J.P., C.A.S., and D.N.B. performed research; L.M., J.I.B., C.A.E., G.E., R.D.P., B.J.D., S.B., and D.N.B. contributed new reagents/analytic tools; L.M., J.I.B., S.M., J.O., L.J.Z., J.D.J., M.R.G., and D.N.B. analyzed data; L.M., J.I.B., M.R.G., and D.N.B. wrote the paper; J.I.B. generated the mutant library and performed deep-sequencing; J.P. performed structure simulation; C.A.E. provided BCR-ABL mutation information in CML samples; G.E. provided BCR-ABL mutation information in CML samples; R.D.P. provided BCR-ABL mutation information in CML samples; B.J.D. provided BCR-ABL mutation information in CML samples; S.B. provided BCR-ABL mutation information in CML samples; and C.A.S. performed structure simulation.

Reviewers: Y.B., Rutgers University; and M.T., University of Toronto.

The authors declare no conflict of interest.



[1]Present address: Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85281.

[2]To whom correspondence may be addressed. Email: Michael.Green@umassmed.edu or Dan.Bolon@umassmed.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1708268114/-/DCSupplemental.

## Results

**An Optimized CRISPR-Cas9–Based Strategy Efficiently Integrates Barcoded *Bcr-Abl1* Libraries into a Consistent Genomic Location.** We carried out proof-of-principle studies of our mutagenesis screening strategy on BCR-ABL1 because: (*i*) clinical resistance to tyrosine kinase inhibitors (TKIs) is frequently associated with mutations in BCR-ABL1 (8, 9) and (*ii*) the clinical drug-resistant mutation spectrum for BCR-ABL1 is well known, allowing us to evaluate the power of this approach (10, 11). The majority of TKI-resistant BCR-ABL1 mutations are observed within the kinase domain (8), leading us to select this region for targeted saturation mutagenesis.

We first stably transduced a nonfunctional BCR-ABL1 mutant lacking the kinase domain (BCR-ABL1ΔKD) into murine Ba/F3 cells, which are conditionally dependent on BCR-ABL1 function in the absence of interleukin 3 (IL-3) (12). Site-specific introduction of a wild-type ABL1 kinase domain (ABL1KD) using CRISPR-Cas9 generates full-length BCR-ABL1, promotes IL-3–independent growth, and induces expression of a GFP reporter, allowing us to screen a library of ABL1KD variants (Fig. 1*A* and *SI Appendix*, Fig. S1). To systematically sample hundreds of BCR-ABL1 mutations in the same experiment, we extensively optimized the CRISPR-Cas9 integration strategy using flow cytometry with a donor template containing DsRed in place of the ABL1KD (*SI Appendix*, *SI Materials and Methods* and Fig. S2). The efficiency of homology-directed repair (HDR) was increased from initial levels by sequential optimization of the sequences of synthetic guide RNAs (sgRNAs) and donor templates (step 1), homology arm length (step 2), clonal cell line (step 3), and amount of sgRNA and donor template (step 4) (*SI Appendix*, *SI Materials and Methods* and Fig. S2*D*), to achieve an HDR efficiency of ∼8% (Fig. 1*B*), representing the successful genetic modification of 1–2 × $10^5$ cells per reaction. For 10-fold coverage of mutant libraries, this efficiency is sufficient to screen up to 20,000 mutations in a single reaction. As part of the HDR optimization process, we identified a clone possessing the highest HDR activity (*SI Appendix*, Fig. S2*D*). When rescued with wild-type ABL1KD, the clone produced a TKI treatment response that resembled that of human BCR-ABL$^+$ cells (*SI Appendix*, Fig. S3*A*). In this clone, we observed that BCR-ABL1ΔKD was integrated at a single genomic locus within transcriptionally active chromatin (*SI Appendix*, *SI Materials and Methods* and Fig. S3*B*). This clone was used in all subsequent experiments.

We first generated a saturated mutagenesis library at the so-called gatekeeper position (T315), whose mutation to isoleucine is known to confer resistance to the first generation TKI imatinib (9). To facilitate multiple independent measurements of each amino acid change in the ABL1KD mutant libraries, we introduced a barcoded intron into the constant region downstream of the ABL1KD, with an average of 25 barcodes associated with each mutant (Fig. 1*A* and *SI Appendix*, Fig. S4*A*). We designed the barcodes to avoid sequences that would disturb intron splicing and protein expression (*SI Appendix*, *SI Materials and Methods* and Fig. S4). The frequency of mutations that we observed integrated into the genome closely matched the frequency in the donor library at both the amino acid and codon levels (Fig. 1 *C* and *D*), indicating that integration was neither strongly biased by the mutations nor a bottleneck.

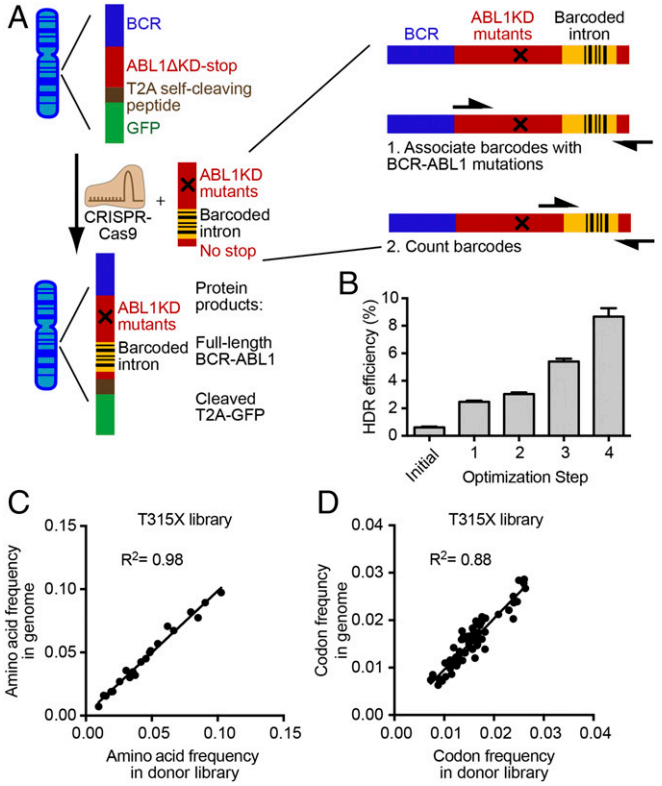



**Fig. 1.** An optimized CRISPR-Cas9–based strategy efficiently integrates barcoded BCR-ABL1 libraries into a consistent genomic location. (*A*) Schematic of the CRISPR-Cas9–based strategy to introduce barcoded libraries of ABL1KD mutations into the genome. (*B*) HDR efficiency following optimization of the sgRNA and donor template sequences (step 1), homology arm length (step 2), clonal cell line (step 3), and the amount of guide RNA and donor template (step 4). The initial HDR efficiency before optimization is shown. Error bars represent the SD of three independent experiments. (*C* and *D*) Correlation of amino acid (*C*) and codon (*D*) frequencies of the library of T315X mutants between the donor library and those integrated into the genome. The plots show all 20 amino acids (including wild type) and stop codon (*C*), and all 64 codons (*D*).

**Fitness of T315X Mutations Under Various Selection Pressures.** Cells transfected with the T315X library were grown in the presence of IL-3 for the first 3 d, then washed and expanded in IL-3–free medium for 3 additional days. On day 6, cells were treated with or without imatinib for an additional 3 d in the absence of IL-3 (Fig. 2*A*). Based on deep sequencing of barcodes, we determined the frequency of each amino acid at position 315 before IL-3 withdrawal (day 3) and at the end of the experiment (day 9); mutations that compromise cell growth will be underrepresented at the later time points, whereas mutations that enhance cell growth will be relatively overrepresented.

We first determined the effects of mutations on growth in the absence of IL-3 and imatinib (i.e., growth effects). Most amino acid changes at position 315 caused a growth defect, whereas five mutations (L, M, I, V, and E) increased IL-3–independent growth (Fig. 2*B*, *Top*). Supporting these observations, mutations corresponding to T315I and T315M have been reported to increase kinase activity in c-ABL1 and/or c-Src (13). In replicate growth competitions, we observed a strong correlation ($R^2$ = 0.91) between estimates of BCR-ABL1 function (Fig. 2*C*).

Similar to traditional $IC_{50}$ measurements, the relative growth of mutants in the presence or absence of imatinib was compared to estimate the effects of mutations on drug binding (i.e., inhibitor effects). Based on this analysis, all mutations at position 315 compromised drug binding (Fig. 2*B*, *Middle*). Traditional $IC_{50}$s were determined for a panel of individual mutations (L, M, I, E, and C; *SI Appendix*, Fig. S5*A*) and demonstrated consistency with our bulk analyses (Fig. 2*D*).

In principle, the combined effects of mutations on BCR-ABL1 function and drug binding will be under selection in patients. In our experiments, we observed strong combined effects for six amino acid changes (L, M, I, V, Q, and E) (Fig. 2*B*, *Bottom*). Five of them (L, M, I, Q, and E) exhibited increasingly adaptive responses with elevated imatinib concentration (Fig. 3*A*, *Top*).

We also examined selection in the presence of two additional TKIs, dasatinib (*SI Appendix*, Fig. S5*B*) and ponatinib (Fig. 3*A*, *Bottom*), the latter of which was designed to counteract the T315I mutation (14). Consistent with its design, ponatinib inhibited T315I. However, the four remaining mutations at position 315 (L, M, Q, and E) appeared increasingly adaptive with higher ponatinib concentrations. Again, traditional $IC_{50}$s for a panel of individual

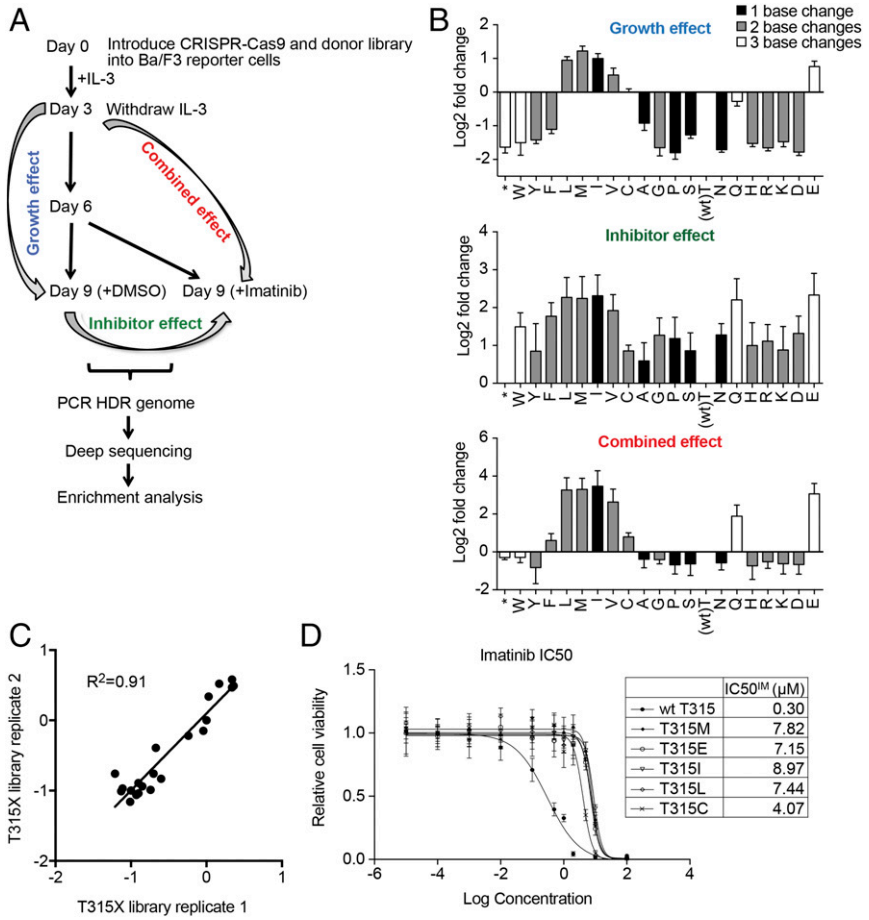



**Fig. 2.** Fitness of T315 mutations under various selection pressures. (*A*) Flowchart showing the experimental pipeline for analysis of mutant enrichment following various growth/inhibitor selection conditions. (*B*) Enrichment analysis of T315X mutants for growth effect [day 9 (−IL-3, DMSO) versus day 3 (+IL-3)], inhibitor effect [day 9 (−IL-3, imatinib) versus day 9 (−IL-3, DMSO)] and combined effect [day 9 (−IL-3, imatinib) versus day 3 (+IL-3)]. Mutations that are accessible through one, two, or three base changes are indicated. Error bars represent SE from three independent experiments. (*C*) Correlation of growth effect of all amino acid mutants between two replicates of the T315X library. (*D*) $IC_{50}$ values of selected T315X mutants for imatinib.

mutations (L, M, I, E, and C) demonstrated consistency with our bulk analyses (*SI Appendix*, Fig. S5*C*).

To explore how these amino acid changes might evolve, we examined the number of nucleotide substitutions required to achieve the mutation and the potential single base substitution pathways (Fig. 3*B*). Of the six amino acid changes at position 315 that are adaptive to imatinib (L, M, I, V, Q, and E), only T315I is accessible by a single base substitution, the most accessible type of mutation (15). In principle, the prevalence of T315I in patient isolates (11) could be due to mutational accessibility and/or adaptive potential. Our data indicate similar adaptive potential to imatinib for L, M, I, and E, with L, M, and E having slightly greater adaptive potential than I at high imatinib concentrations (Fig. 3*A*, *Top*). These observations suggest that mutational accessibility has a strong influence on the evolution of drug resistance in BCR-ABL1. At high concentrations of imatinib or ponatinib, T315M and T315E were the most adaptive (Fig. 3*A* and *SI Appendix*, Fig. S5*C*). T315M is accessible by a fitness-increasing path, while T315E cannot be accessed by single nucleotide substitutions without going through a maladaptive intermediate (Fig. 3*B*). The observed advantage of T315M relative to T315I is greater under ponatinib treatment compared with imatinib (Fig. 3*A*), indicating that a mutational pathway from T to I to M may be likely in patients who have transitioned from imatinib to ponatinib therapy. Consistent with this hypothesis, T315M has recently been reported to evolve from T315I during ponatinib treatment of a patient who previously failed imatinib therapy (16).

**Evolutionary Adaptation of BCR-ABL1 Mutations.** Based on this predictive potential, we examined a larger region that flanks the gatekeeper residue, positions 311–319 of BCR-ABL1. While most mutations were deleterious, a number of mutations at positions 311, 315, 317, and 319 increased adaptive potential of BCR-ABL1 (Fig. 4*A*). To assess the reliability of this higher-throughput experiment, we compared independent replicates (Fig. 4*B*) and also compared the results for position 315 within the 311–319X library to the single position T315X experiment (Fig. 4*C*). The positive correlation between these experiments ($R^2 = 0.66$ and $R^2 = 0.95$, respectively) indicates that both experiments provide accurate measurements of selection pressures acting on these mutations, although reproducibility varied for individual positions (*SI Appendix*, Fig. S6).

The growth effect analysis of Fig. 4*A* revealed that different positions exhibited strikingly distinct patterns of functional sensitivity to mutation. For instance, at position 316 every amino acid substitution caused a strong growth defect, whereas at position 319 all amino acid changes exhibited minimal impacts on function. These sensitivities were generally consistent with inferences from structure. For example, E316, which is sensitive to amino acid changes, makes multiple hydrogen bonds with Q300 and K378 (*SI Appendix*, Fig. S7*A*); notably, E316 and K378 are highly conserved across 94 tyrosine kinases (*SI Appendix*, Fig. S7 *B* and *C*). By contrast, the side chain of T319 is oriented toward solvent such that amino acid changes can be tolerated without directly disrupting structure (*SI Appendix*, Fig. S7*A*).

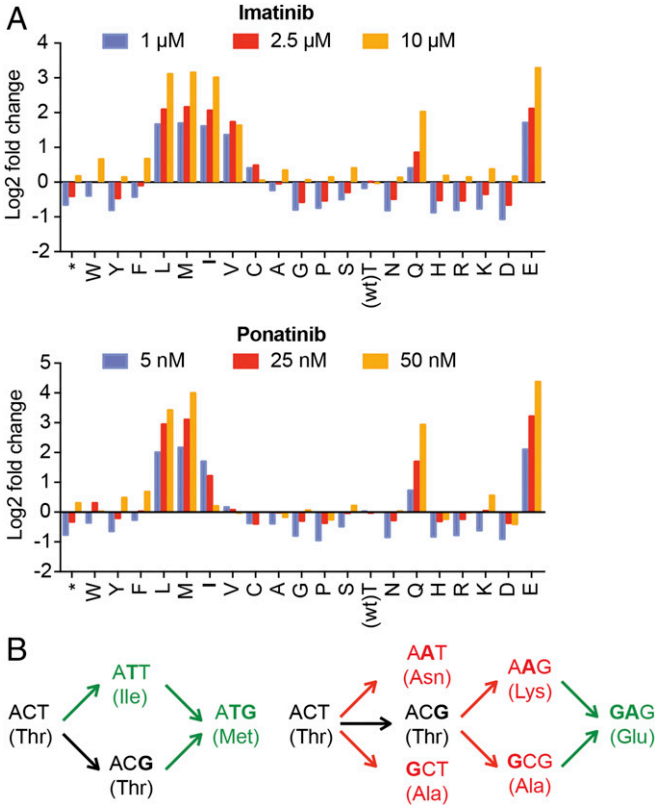



**Fig. 3.** Increased adaptation to TKI therapy accessible by single base substitution. (*A*) Enrichment analysis of T315X mutants for combined effect at increasing concentrations of imatinib (*Top*) or ponatinib (*Bottom*). (*B*) Single base mutational pathway from Thr to Met or Glu. Mutations with increased fitness are marked in green, decreased fitness in red, near neutral mutations in black.

**BIG EMPIRIC Accurately Predicts Clinical Prevalence of BCR-ABL1 Mutations.** We sought to explore whether BIG EMPIRIC is capable of predicting the clinical prevalence of each mutation as assessed from a collection of BCR-ABL mutations in ~1,400 imatinib-resistant patient samples (Fig. 5*A* and *SI Appendix*, Table S1). The mutations with the strongest combined effects were located at position 315 (Fig. 5*B*), which is consistent with mutations at this position evolving as the most commonly observed imatinib-resistant mutations in clinical samples (11).

Encouraged by this observation, we investigated whether the combined effects together with mutational probabilities could accurately predict clinical observations of imatinib resistance. Indeed, single nucleotide mutations with the greatest combined effects in our screen were also observed with the highest frequency in clinical samples (Fig. 5 *C* and *D*). For these single nucleotide mutations, we calculated a prediction score based on the observed combined effects and mutational probabilities, which correlated very well with their observed prevalence in patients (Fig. 5*E*). Although the strength of this correlation was primarily influenced by the two most clinically prevalent mutations (T315I and T317L), it is encouraging that the prediction score accurately estimates the frequency of all clinically observed mutations analyzed in the 311–319 region. Furthermore, the combined effects provided more accurate predictions of clinical prevalence than either growth effect or inhibitor effect alone (*SI Appendix*, Fig. S8 *A* and *B*), indicating that both properties contribute to the evolution of drug resistance.

Protein folding stability can be a dominant force in evolution (17), and we therefore investigated whether the estimated impacts of mutations on protein folding stability, calculated using FoldX (18), might correlate with the experimental impacts of mutations in BCR-ABL1. The analysis revealed weak correlations ($R^2 < 0.1$) between estimates of folding stability in either the active or inactive states of BCR-ABL1 and our measurements of the combined effects (*SI Appendix*, Fig. S8 *C* and *D*). Despite the poor correlation with combined effects, we also examined whether protein folding stability estimates combined with mutational probabilities might predict clinical prevalence. Predictions based on active and inactive conformations also correlated weakly ($R^2 < 0.2$) with clinical prevalence (*SI Appendix*, Fig. S8 *E* and *F*). Thus, simple estimates of the impacts of mutations on BCR-ABL1 folding stability were unable to accurately predict the experimental or clinical impacts of drug resistance mutations. Molecular dynamic simulations of a handful of individual mutants (*SI Appendix*, Fig. S8*G*) also suggest that gross structural perturbations fail to explain their observed function.

## Discussion

In this report, we have described an improved saturation mutagenesis method with high reproducibility, called BIG EMPIRIC, and demonstrated its capability of quickly interpreting the functional consequence of hundreds of mutations. BIG EMPIRIC has several advantages over previously developed high-throughput mutagenesis methods (3, 4, 19–21). First, BIG EMPIRIC systematically generates all point mutations at similar frequencies, in contrast to random mutagenesis, which tends to generate an unbalanced library that can leave out mutations (3, 4, 19, 20). Second, in BIG EMPIRIC, mutations are analyzed at a single genomic location, which should produce consistent expression compared with approaches utilizing random lentiviral integration of mutant genes. Third, BIG EMPIRIC is a flexible strategy that can be easily adapted to many other oncogenes, without the limitation of conducting screens at unique endogenous genetic contexts, such as short exons or splicing junctions (21). Fourth, the intron-based barcoding strategy of BIG EMPIRIC can facilitate analyses of larger regions of genes in the same experiment and provides many internal reproducibility checks for quality-control purposes.

Like many other technologies, however, BIG EMPIRIC is not without its limitations. First, the saturated library is limited to a 10 amino acid window. However, this shortcoming could be circumvented by combining a few libraries for each reaction

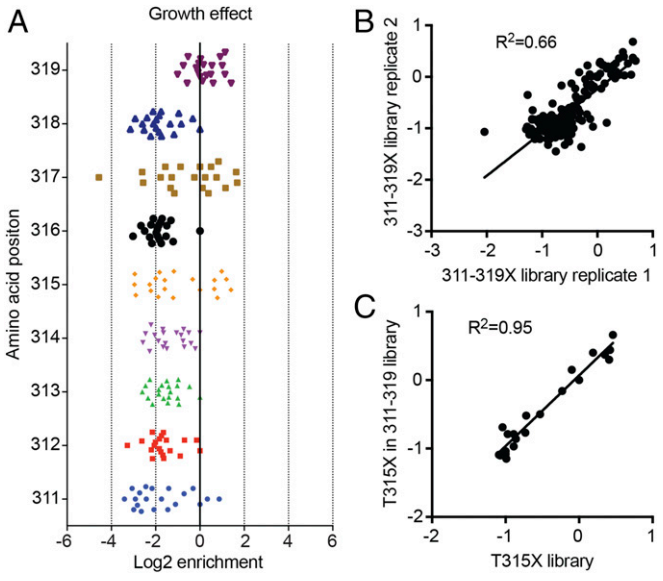
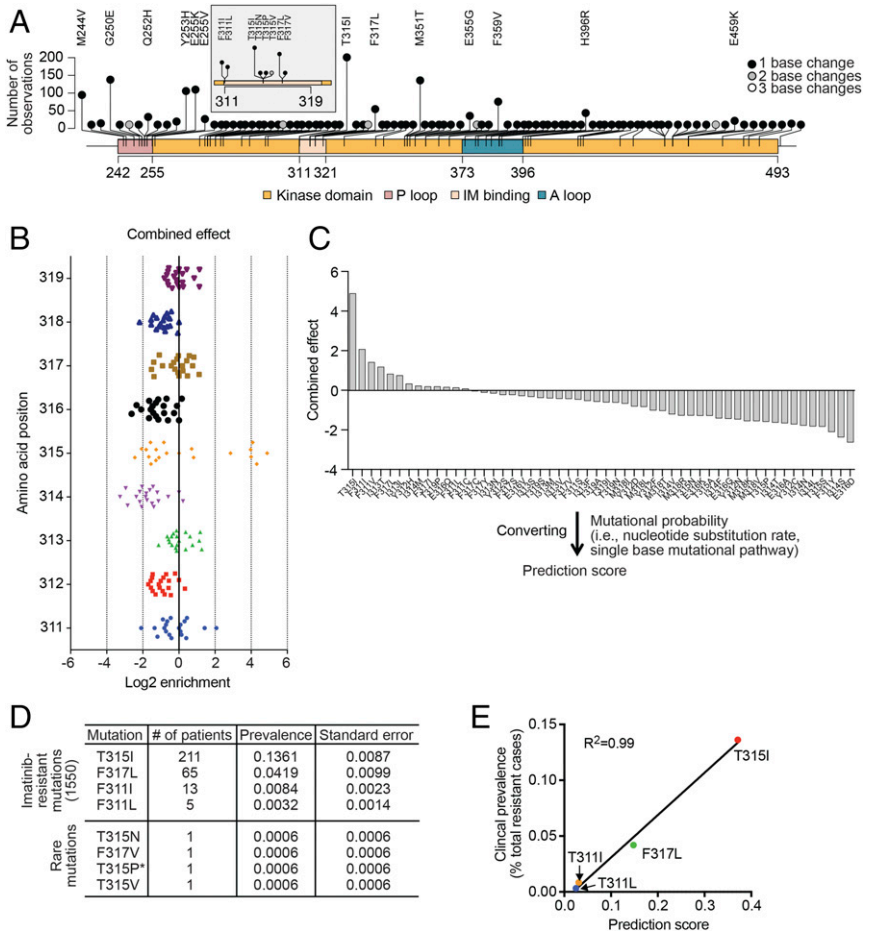



**Fig. 4.** Evolutionary adaptation of BCR-ABL1 mutations. (*A*) Growth effect of all possible mutations at each amino acid position in the 311–319 region. Each group comprises 20 dots, representing all possible amino acid variants; the wild-type amino acid was set to 0. (*B*) Correlation of growth effect of all amino acid mutations between two replicates of the 311–319X library. (*C*) Correlation of growth effect of all T315X mutants between the T315X library and 311–319X library.




| | Mutation | # of patients | Prevalence | Standard error |
|---|---|---|---|---|
| Imatinib-resistant mutations (1550) | T315I | 211 | 0.1361 | 0.0087 |
| | F317L | 65 | 0.0419 | 0.0099 |
| | F311I | 13 | 0.0084 | 0.0023 |
| | F311L | 5 | 0.0032 | 0.0014 |
| Rare mutations | T315N | 1 | 0.0006 | 0.0006 |
| | F317V | 1 | 0.0006 | 0.0006 |
| | T315P* | 1 | 0.0006 | 0.0006 |
| | T315V | 1 | 0.0006 | 0.0006 |

**Fig. 5.** BIG EMPIRIC accurately predicts clinical prevalence of BCR-ABL1 mutations. (*A*) Lollipop plot showing the distribution of all clinically identified BCR-ABL mutations from a collation of ∼1,400 imatinib-resistant patient samples. Mutations that are accessible through one, two, or three base changes are indicated. Data were collated from the Branford laboratory (10), COSMIC database, and Oregon Health & Science University. (*B*) Combined effects of all possible mutations at each amino acid position in the 311–319 region. Each group comprises 20 dots, representing all possible amino acid variants; the wild-type amino acid was set to 0. (*C*) Ranking of the combined effect scores for amino acid substitutions achievable through a single-base change within the 311–319 region. (*D*) Clinically identified amino acid mutations in the 311–319 region. The asterisk indicates a previously unreported mutation with confirmed imatinib resistance in the current study. The SE of the proportion of each amino acid change observed in the clinical samples is shown. (*E*) Correlation between prediction score (based on the observed combined effects and mutational probabilities) and clinical prevalence for four mutations.

without surpassing the screening capacity, and multiple reactions could be performed to cover an entire ORF. Second, finding a suitable clone with ideal HDR potential can be laborious. Of note, the efficient genomic locus that we identified here may work well in future screens of other oncogenes in Ba/F3 cells.

We used BIG EMPIRIC to generate a prediction score that correlated highly with the clinical prevalence of identified BCR-ABL1 mutations, suggesting that it would be a useful method for predicting clinical drug-resistant mutations in other less well-studied drug targets. Notably, the observation that only a few types of T315 mutations have been clinically observed (16) indicates that identifying mutants with altered drug binding alone may be misleading with regard to clinical relevance. Accordingly, our analysis revealed that an accurate assessment of the drug response of each mutant together with its adaptive potential lead to better prediction of its clinical relevance. Our observations highlight the utility of observing growth effects to understand the evolution of drug resistance in cancer.

In this study, we focused on amino acid changes accessible by single nucleotide mutations, as they are by far the most frequently observed in drug resistant cancers (15). In addition, all of the clinically observed imatinib-resistant amino acid changes in the regions of BCR-ABL1 that we explored experimentally (Fig. 5*E*) are accessible by single nucleotide mutations. In our experiments, single nucleotide mutations (Fig. 5*C*) had similar combined effects as multiple nucleotide mutations (*SI Appendix*, Fig. S8 *H* and *I*). One explanation for the prevalence of single nucleotide mutations in imatinib-resistant cancer is that double nucleotide mutations do not provide a further benefit. By contrast, multiple nucleotide substitutions appear important for the evolution of ponatinib resistance in BCR-ABL1 (16). However, there is insufficient clinical data at this time to examine the potential accuracy of mutational pathway predictions involving multiple base substitutions.

The responses of different cancers to targeted therapy can vary tremendously with regard to mechanism and timescale (2, 22). Our results suggest that differences in emergence of resistance are likely due to differences in readily available mutational pathways, as well as the impacts of potential mutants on growth and drug binding. In this regard, BIG EMPIRIC provides a promising approach to help tease these factors apart for different oncogenes.

The mutations observed to cause resistance in patients are often not clear until drugs are utilized clinically. Research on resistance mechanisms has the capability to inform drug development (23), especially when all of the relevant structures are not known. Within the conventional drug development paradigm, it could take up to 10 y before gathering enough information to judge the clinical efficacy of a drug and its sensitivity to mutations (24, 25). BIG EMPIRIC offers an alternative strategy to improve drug design

through multiple rounds of predicted clinical response within months at early drug development stage, which potentially helps avoid clinical resistance later (*SI Appendix*, Fig. S9).

## Materials and Methods

**Cell Lines and Cell Culture.** Ba/F3 reporter cell line generation, including reporter plasmid construction, HDR optimization, and verification of integration of the reporter, is described in *SI Appendix, SI Materials and Methods*. All Ba/F3 cell lines were maintained in RPMI medium 1640 containing 10% FBS, 4 mM L-glutamine, 100 units/mL penicillin, and 100 μg/mL streptomycin. Parental Ba/F3 cell culture was supplemented with 10 ng/mL IL-3, and clonal Ba/F3 cells containing doxycycline-inducible BCR-ABL1 were supplemented with 1 μg/mL of doxycycline (cat. no. D9891; Sigma) unless otherwise stated.

**Library Introduction and Drug Treatment.** Construction of the T315X and 311–319X donor libraries are described in *SI Appendix, SI Materials and Methods*. Before transfection of the library, on day −4, $1 \times 10^6$ Ba/F3 reporter cells were cultured initially in a T75 flask for 3 d in the presence of IL-3 but without doxycycline. At day −1, cells were collected, pelleted, and resuspended in 15 mL of IL-3–containing fresh medium for every $1 \times 10^7$ cells, then cultured for an additional 24 h. At day 0, cells were collected for electroporation and transfected with the library using the optimal parameters defined in the HDR optimization section (*SI Appendix, SI Materials and Methods*).

On day 1, cells were transferred from a 12-well plate into a T25 flask with fresh IL-3–containing medium, and doxycycline (1 μg/mL) was added. On day 3, cells were pelleted and washed twice in IL-3–free medium. An aliquot of cells ($\sim 1–2 \times 10^6$) was saved to serve as a control sample. The remaining cells were cultured in IL-3–free and doxycycline-containing medium in a T75 flask for an additional 3 d, unless otherwise stated. On day 6, cells were collected and split, and half were treated with DMSO and the other half with a TKI [2.5 μM imatinib (cat. no. CT-IM001, ChemieTek), 20 nM dasatinib (cat. no. CT-DS001, ChemieTek), or 10 nM ponatinib (cat. no. P-7022, LC Labs)] and grown for an additional 3 d in the absence of IL-3. On day 9, all cells were pelleted and frozen at −80 °C for future genomic DNA extraction and nested PCR, as described in *SI Appendix, SI Materials and Methods*.

**Fitness and Drug Resistance Analysis.** Deep sequencing results were processed by custom perl scripts. First, sequences with Phred scores >10 containing the constant region of the PCR products were extracted from the fastq file. Extracted sequences were then parsed based on the sample ID added in the second round of PCR; IDs correspond to the growth condition the PCR product was generated from. Next, counts for each unique barcode were generated for each condition/sample ID. Barcodes were then decoded using the subassembly file described in the "T315X and T311-319X HDR Donor Library Construction" section (*SI Appendix, SI Materials and Methods*). Finally, counts for barcodes corresponding to the same codon or amino acid mutation were merged.

The mutation frequencies for each condition were analyzed in three ways. First, the ratio of the amino acid frequency at "day 9, −IL-3, DMSO" versus "day 3, +IL-3" was log2 transformed and normalized to the wild-type synonyms; this analysis, termed the "growth effect," highlights the effect each mutation has solely on cell growth in the absence of inhibitor. Second, the ratio of the amino acid frequency at "day 9, −IL-3, inhibitor" versus "day 9, −IL-3, DMSO" was log2 transformed and normalized to wild-type synonyms; this analysis, termed the "inhibitor effect," looks at the effect each mutation has on inhibitor resistance. Third, the ratio of the amino acid frequency at "day 9, −IL-3, inhibitor" versus "day 3, +IL-3" was log2 transformed and normalized to the wild-type synonyms; we refer to this analysis as the "combined effects" because influences of mutations on both growth without drug as well as the affinity of inhibitor contribute to this ratio. The combined effects are static for each pair of amino acids.

To generate a prediction score for the natural evolution of mutations, we used our experimental data for amino acid changes to estimate combined effects together with previously determined mutational probabilities. For a specific amino acid change, we considered the parental codon and all single base mutations to the target amino acid. For example for the change of ATA (Ile) to Leu, there are two available single base mutations (to TTA, or CTA), and we estimated the net probability as the sum of these two paths. The prediction score for each amino acid change in Fig. 5*E* was calculated using the following equation:

$$\sum_1^n \text{Score of combined effect for amino acid subsitution} \times \text{single nucleotide sustitution rate}$$

where $n$ = number of single base mutational paths for each amino acid pair; score of combined effect for amino acid substitution is defined as the ratio of the "day 9, −IL-3, inhibitor" to the "day 3, +IL-3," which is log2 transformed and normalized to wild-type synonyms; and nucleotide substitution rate is defined as the rate for a single nucleotide substitution at a particular codon position that leads to the corresponding amino acid change (26).

**ACKNOWLEDGMENTS.** We thank Sara Deibler for providing editorial assistance and the University of Massachusetts Medical School Deep Sequencing Core Facility for deep sequencing. This work was supported by NIH Grant R01CA163926 (to M.R.G.), and NIH Grant R21CA175664 (to D.N.B.). M.R.G. is also an investigator of the Howard Hughes Medical Institute.

1. Gottesman MM (2002) Mechanisms of cancer drug resistance. *Annu Rev Med* 53:615–627.
2. Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG (2013) Cancer drug resistance: An evolving paradigm. *Nat Rev Cancer* 13:714–726.
3. Azam M, Latek RR, Daley GQ (2003) Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR-ABL. *Cell* 112:831–843.
4. Ma Y, et al. (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat Methods* 13:1029–1035.
5. Wagenaar TR, et al. (2014) Resistance to vemurafenib resulting from a novel mutation in the BRAFV600E kinase domain. *Pigment Cell Melanoma Res* 27:124–133.
6. Fowler DM, Fields S (2014) Deep mutational scanning: A new style of protein science. *Nat Methods* 11:801–807.
7. Hietpas RT, Jensen JD, Bolon DN (2011) Experimental illumination of a fitness landscape. *Proc Natl Acad Sci USA* 108:7896–7901.
8. Soverini S, et al. (2011) BCR-ABL kinase domain mutation analysis in chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: Recommendations from an expert panel on behalf of European LeukemiaNet. *Blood* 118:1208–1215.
9. Gorre ME, et al. (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* 293:876–880.
10. Branford S, Melo JV, Hughes TP (2009) Selecting optimal second-line tyrosine kinase inhibitor therapy for chronic myeloid leukemia patients after imatinib failure: Does the BCR-ABL mutation status really matter? *Blood* 114:5426–5435.
11. Jones D, et al. (2009) Laboratory practice guidelines for detecting and reporting BCR-ABL drug resistance mutations in chronic myelogenous leukemia and acute lymphoblastic leukemia: A report of the Association for Molecular Pathology. *J Mol Diagn* 11:4–11.
12. Daley GQ, Baltimore D (1988) Transformation of an interleukin 3-dependent hematopoietic cell line by the chronic myelogenous leukemia-specific P210bcr/abl protein. *Proc Natl Acad Sci USA* 85:9312–9316.
13. Azam M, Seeliger MA, Gray NS, Kuriyan J, Daley GQ (2008) Activation of tyrosine kinases by mutation of the gatekeeper threonine. *Nat Struct Mol Biol* 15:1109–1118.
14. O'Hare T, 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.
15. Vogelstein B, et al. (2013) Cancer genome landscapes. *Science* 339:1546–1558.
16. Zabriskie MS, et al. (2014) BCR-ABL1 compound mutations combining key kinase domain positions confer clinical resistance to ponatinib in Ph chromosome-positive leukemia. *Cancer Cell* 26:428–442.
17. Wylie CS, Shakhnovich EI (2011) A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc Natl Acad Sci USA* 108:9916–9921.
18. Schymkowitz J, et al. (2005) The FoldX web server: An online force field. *Nucleic Acids Res* 33:W382–W388.
19. Ray A, Cowan-Jacob SW, Manley PW, Mestan J, Griffin JD (2007) Identification of BCR-ABL point mutations conferring resistance to the Abl kinase inhibitor AMN107 (nilotinib) by a random mutagenesis study. *Blood* 109:5011–5015.
20. Hess GT, et al. (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Methods* 13:1036–1042.
21. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J (2014) Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* 513:120–123.
22. Gottesman MM, Lavi O, Hall MD, Gillet JP (2016) Toward a better understanding of the complexity of cancer drug resistance. *Annu Rev Pharmacol Toxicol* 56:85–102.
23. Nalam MN, et al. (2013) Substrate envelope-designed potent HIV-1 protease inhibitors to avoid drug resistance. *Chem Biol* 20:1116–1124.
24. Capdeville R, Buchdunger E, Zimmermann J, Matter A (2002) Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat Rev Drug Discov* 1:493–502.
25. Hajduk PJ, Greer J (2007) A decade of fragment-based drug design: Strategic advances and lessons learned. *Nat Rev Drug Discov* 6:211–219.
26. Zhang Z, Gerstein M (2003) Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. *Nucleic Acids Res* 31:5338–5348.