Cell Chemical Biology
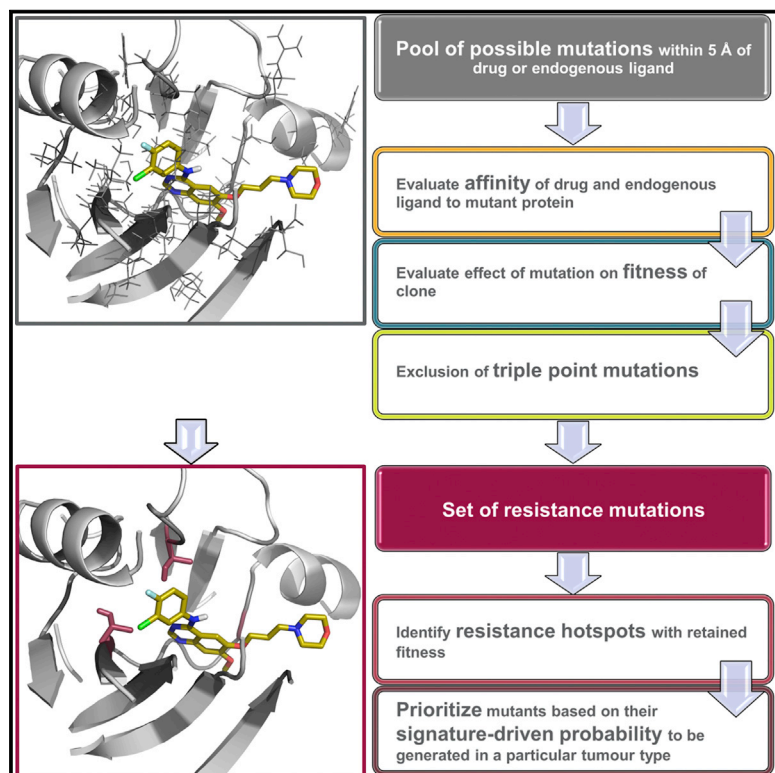
Combining Mutational Signatures, Clonal Fitness, and Drug Affinity to Define Drug-Specific Resistance Mutations in Cancer

Graphical Abstract

Authors

Teresa Kaserer, Julian Blagg

Correspondence

teresa.kaserer@icr.ac.uk (T.K.),
julian.blagg@icr.ac.uk (J.B.)

In Brief

Resistance to targeted cancer drugs is a major challenge in therapy. Kaserer and Blagg report a computational approach to prospectively identify and prioritize clinically relevant drug-resistant mutations; it facilitates improved patient monitoring for the emergence of resistance and the timely discovery of durable treatment options.

Highlights

- Drug affinity, clonal fitness, and mutation signatures define resistant mutations

- We report a computational work flow to identify drug-resistant hotspot mutations

- A comprehensive ERK2 resistance mutation dataset is used to validate the work flow

- Clinically relevant, drug-resistant mutations are identified

CellPress

Combining Mutational Signatures, Clonal Fitness, and Drug Affinity to Define Drug-Specific Resistance Mutations in Cancer

Teresa Kaserer[1,*] and Julian Blagg[1,2,*]
[1]Cancer Research UK Cancer Therapeutics Unit, The Institute of Cancer Research, London SM2 5NG, UK
[2]Lead Contact
*Correspondence: teresa.kaserer@icr.ac.uk (T.K.), julian.blagg@icr.ac.uk (J.B.)
https://doi.org/10.1016/j.chembiol.2018.07.013

SUMMARY

The emergence of mutations that confer resistance to molecularly targeted therapeutics is dependent upon the effect of each mutation on drug affinity for the target protein, the clonal fitness of cells harboring the mutation, and the probability that each variant can be generated by DNA codon base mutation. We present a computational workflow that combines these three factors to identify mutations likely to arise upon drug treatment in a particular tumor type. The Osprey-based workflow is validated using a comprehensive dataset of ERK2 mutations and is applied to small-molecule drugs and/or therapeutic antibodies targeting KIT, EGFR, Abl, and ALK. We identify major clinically observed drug-resistant mutations for drug-target pairs and highlight the potential to prospectively identify probable drug resistance mutations.

INTRODUCTION

Although targeted cancer therapies, for example, against kinases, hormone receptors, or hormone-synthesizing enzymes, have shown clinical success, many patients develop resistance to treatment and subsequently relapse. Second- and third-generation drugs are being developed to target these resistant mutants; however, there is a significant time span between the detection of clinically validated resistance mutations and the availability of suitably targeted treatment options. Early identification of drug-specific mutations is therefore critical and the aim of this study.

Several mechanisms underlying resistance to targeted drugs have been described (Holohan et al., 2013), including mutations directly affecting the drug target. Such mutations may, for example, increase affinity for the endogenous co-factor ATP, thereby decreasing the relative affinity of an ATP-competitive drug. Mutations within the binding site may also alter drug-protein interactions and directly interfere with drug binding (Barouch-Bentov and Sauer, 2011).

Several computational studies have investigated the impact of protein mutations on drug efficacy (Kamasani et al., 2017; Mar-

tínez-Jiménez et al., 2017; Paul et al., 2016; Reeve et al., 2015), mainly in an antiviral or antibacterial context (Frey et al., 2010; Hosseini et al., 2016; Fowler et al., 2018). However, rarely do these studies prospectively identify drug resistance mutations (Frey et al., 2010; Paul et al., 2016; Reeve et al., 2015). In predicting mutations that render *Staphylococcus aureus* resistant to an antifolate antibiotic, Reeve et al. (2015) evaluated the likely effect of possible mutations on both binding of the inhibitor and on binding of the endogenous ligand — an important aspect since any mutation that significantly abrogates the native activity of the wild-type (WT) protein is unlikely to survive selective evolutionary pressure (Gil and Rodriguez, 2016; Sprouffske et al., 2012; Pandurangan et al., 2017). However, Reeve et al. do not consider the likelihood of whether each mutation can be formed in bacteria.

In cancer, the mutation landscape of a tumor can be characterized by the mutational signatures operating in a particular cancer type (Alexandrov et al., 2013). These signatures describe the probability of a specific base exchange within a defined trinucleotide context. Some of these signatures have been associated with known mutagenic processes, such as UV irradiation or aging, while the mechanism of others still remains elusive (Alexandrov et al., 2013). These mutagenic processes can generate a single clone harboring the disease-causing "driver mutation," which ultimately leads to the development of cancer (Greaves and Maley, 2012). In addition, non-transforming somatic mutations, so-called passenger mutations, are randomly created. While not oncogenic per se, passenger mutations can provide the substrate for an evolutionary advantage throughout cancer progression, for example, under the selective pressure of a targeted molecular therapy, leading to drug resistance. Known drug resistance mutations have not only been detected in treatment-naive patients (Inukai et al., 2006; Roche-Lestienne et al., 2002), but also in healthy individuals (Gurden et al., 2015). This suggests that small pools of viable treatment-resistant clones can pre-exist in patients and that drug treatment puts a selection pressure on a heterogeneous cancer cell population that selects for resistant sub-clones.

Each drug interacts with its biological target in a unique way, and each protein target mutation will differentially affect diverse classes of drugs. As a consequence, each compound can be expected to exhibit a unique resistance mutation profile. Three factors contribute to the probability and functional impact of a residue change: (1) the probability that the protein mutation can be generated from a DNA mutational signature (signature-driven
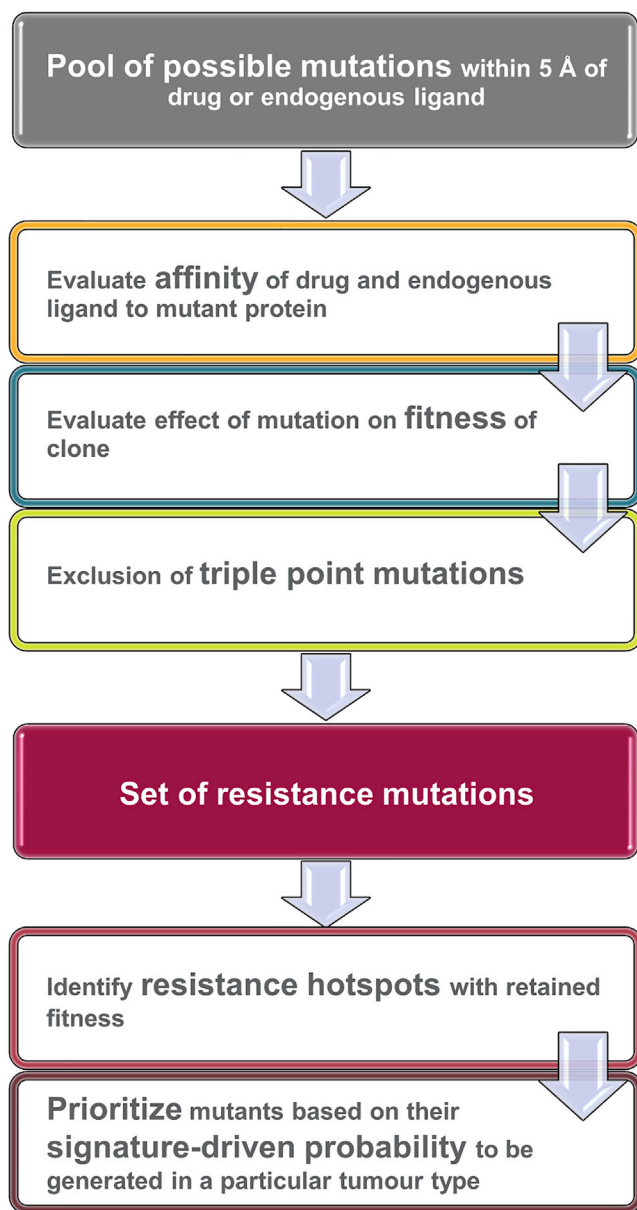
Figure 1. Workflow
Potential mutations are evaluated based on their predicted effect on the affinity of both the drug and endogenous ligand (orange), the fitness of the resultant clone (blue), and the requirement for triple-point mutations to generate a mutant (lime green). Resistance hotspots are identified within the remaining set of resistant mutants; these resistance hotspots are protein residues where multiple amino acid changes are predicted to lead to resistance and which therefore have a high likelihood of functional relevance. Resistant mutations at these hotspots are prioritized based on the probability that they will be generated according to the known DNA mutational signatures operating in a particular cancer type.

probability), (2) whether the mutation maintains protein function and clones harboring the mutation are still viable (fitness), and (3) whether the mutation confers lower drug affinity with respect to the endogenous ligand for the target protein (affinity). Martí-nez-Jiménez et al. (2017) recently reported a workflow classi-

fying potential drug resistance mutations based on Random Forest models and mutation signatures. However, the effect of mutations on the fitness of the clone was not taken into account. In addition, only single-point mutations (SPMs) were considered, despite the notable detection of double-point mutations (DPMs) in cancer patients (Table S1).

We report an *in silico* cascade that sequentially evaluates the probability of generating any mutant within 5 Å of a bound ligand, the clonal fitness of each mutation, and the effect of each mutation on drug affinity in order to systematically and objectively prioritize mutations that are highly likely to arise under drug treatment. Importantly, our workflow classifies the impact of a mutation on drug affinity relative to endogenous ligand and does not rely upon accurate calculation of binding free energies; it also ranks mutations according to their likelihood of being generated in particular cancer type. The workflow (Figure 1 and described in detail below) is validated on a comprehensive benchmark dataset that describes the effect of nearly all possible extracellular signal-regulated kinase 2 (ERK2) missense mutations on sensitivity to the ERK2 inhibitor SCH772984 (Brenan et al., 2016). In addition, we apply the workflow to four well-established cancer targets and evaluate at least two US Food and Drug Administration-approved drugs (small molecules or biologics) per target.

RESULTS

Affinity
In the first step, the software package Osprey (Chen et al., 2009; Gainza et al., 2013) was applied to evaluate the potential impact of protein active site mutations on both drug and endogenous ligand binding using protein crystallographic data.

Osprey uses ensembles of minimized side chain rotamers to calculate a K* score (Chen et al., 2009; Gainza et al., 2013), which estimates ligand binding to the evaluated protein construct. Importantly, by employing ensembles of rotamers, Osprey takes the dynamic nature of protein amino acid side chains into account.

K* scores were computed for the WT protein and all possible amino acid exchanges for both the drug and the endogenous ligand-protein complex with the exception of Pro, which cannot be processed by Osprey. We identify mutants that reduce affinity for the drug in comparison with the endogenous ligand and are therefore more likely to drive a viable resistant clone. Importantly, calculation of absolute drug affinity for WT and mutant proteins is not required. Please refer to the STAR Methods for additional details.

Fitness
Clones harboring a mutant oncoprotein that is unable to bind its endogenous ligand lose their function and are unlikely to persist. All such deleterious mutants with a K* score of 0.0 for the endogenous ligand were discarded.

Exclusion of Triple Point Mutations
Base codons for each WT residue were retrieved from the COSMIC database (Bamford et al., 2004). Amino acid exchanges that would require mutation of all three codon bases were discarded. We hypothesized that three contiguous coding

base mutations are highly unlikely and, to the best of our knowledge, no such patient case has been reported. While the majority of drug resistance mutations arise from SPMs, we also included DPMs (so-called tandem mutations), which represent 5% (7 of 132) of unique drug resistance mutations with defined genetic alterations in the COSMIC database (access date 5 January 2018) and occur with similar frequency to insertions (Table S1).

Resistance Hotspots and Signature-Driven Probability

Functional protein mutants were further triaged according to the number of predicted mutations at a specific residue position. This approach identifies resistance hotspots, where multiple different viable mutants are predicted to cause resistance to the drug under study. While comparing the prevalence of mutations at hotspots in a particular drug-target pair can be used for ranking positions, this method cannot be applied to cross-compare the prevalence of mutations at hotspots between drug-target pairs to infer global inter-target susceptibility to resistance. The top three ranked resistance hotspots were then analyzed in more detail in order to further prioritize amino acid changes at these positions that have a high probability (relP) of being generated in particular tumor types according to the DNA mutational signatures reported by Alexandrov et al. (2013). Each possible resistance hotspot mutation was ranked according to its derived relP and the top-ranked amino acid changes were considered most relevant. Please refer to the STAR Methods for additional details.

ERK2: A Benchmarking Study

In 2016, Brenan and colleagues systematically studied the effect of a wide range of ERK2 genetic alterations on the cellular response to the ERK2 inhibitor SCH772984 (Brenan et al., 2016). They generated nearly every possible missense mutation using multiplexed site-directed mutagenesis. Pooled vectors were used to express ERK2 mutants in A375 human malignant melanoma cells; after drug exposure, the enrichment of resistant variants relative to their abundance in the initial cell population was determined by massively parallel sequencing.

This comprehensive dataset of possible resistant mutants to the ERK2 inhibitor SCH772984 provides a benchmark to evaluate the performance of our workflow. However, as a consequence of the aggressive mutant-generation protocol, this dataset includes protein mutants that require exchange of all three coding bases. Such mutants are excluded in our workflow (see above) and were also removed from the Brenan dataset.

Brenan and colleagues reported 46 experimentally observed resistance mutations within 5 Å of the bound ligand that were not mutations to Pro or triple-point mutations. For SCH772984, 559 mutants covering 31 residues were evaluated by our workflow. The number of true-positive (TP) (mutants predicted to confer resistance that were confirmed by experimental testing), true-negative (mutants predicted to be sensitive that did not confer resistance in experimental testing), false-positive (FP) (mutants predicted to confer resistance that were found to be sensitive in experimental testing), and false-negative (mutants predicted to be sensitive, but conferred resistance in the experimental testing) predictions was calculated (Figure 2A).

Our workflow correctly classifies 84.6% of mutants and identifies almost 40% of the experimentally observed resistance mu-

tations reported by Brenan et al. When considering only residue positions, 80% of experimentally observed mutated residues are discovered. The majority of experimentally evaluated ERK2 mutants remained sensitive to SCH772984 treatment and are correctly identified as sensitive by our approach, contributing to the high overall proportion (84.6%) of correct predictions. The low TP rate for SCH772984 (3.2% of mutants predicted to confer resistance were confirmed by experimental testing) is consistent with the experimental observation that only a small proportion of possible mutants (8.2%) cause drug resistance. The prioritized hotspot residues and their corresponding reported resistance mutations are provided in Table 1. For a detailed list of all predicted mutations please refer to Table S2.

Clinical Case Studies

We studied four protein kinases (KIT, epidermal growth factor receptor [EGFR], breakpoint cluster region-Abelson [Bcr-Abl] kinase, and anaplastic lymphoma kinase [ALK]). In each case protein-ligand crystal structures of approved drugs exemplify multiple generations of compound design; furthermore, as inhibitors of these targets provide the current standard of care for multiple cancer types, clinically observed resistance mutations are documented for most compounds and provide the potential for clinical validation of our computational workflow.

KIT

The first-generation inhibitor imatinib is resistant to the KIT gatekeeper mutation T670I (Antonescu et al., 2005; Tamborini et al., 2004). We investigated 648 possible mutations, all of which lie within 5 Å of either imatinib or ADP; 68 mutations remained after filtering for reduced drug binding, fitness of the clones, and triple-point mutants. For residue T670, eight mutations are predicted to cause resistance to imatinib, making it the first-ranked resistance hotspot; within this set, gatekeeper mutation T670I has the highest relP. Furthermore, the known resistance mutation T670E (Wardelmann et al., 2005) is ranked sixth by relP. Positions C809 and V668 are identified as second- and third-ranked resistance hotspots, respectively. C809R, the mutation with the highest relP, has been reported in a patient diagnosed with myelodysplastic syndrome-derived leukemia (Lorenzo et al., 2006), highlighting its biological relevance; however, similar to all predicted V668 mutations, this residue has not yet been associated with imatinib resistance.

The second-generation inhibitor sunitinib is reported to overcome resistance to the gatekeeper mutation T670I and, consistent with these reports, T670I/E mutations are not predicted to elicit sunitinib resistance by our workflow. Indeed, none of our predicted resistance mutations to sunitinib has yet been associated with clinically observed resistance. Only one resistance mutation has so far been reported for sunitinib in the COSMIC database (Bamford et al., 2004, access date 23 November 2017); this mutation is not within the ATP-binding site.

Ponatinib has been shown to bind KIT, but is not approved for KIT-associated cancer types and clinical data are therefore lacking. Our method predicts resistance to ponatinib due to less-favorable interactions with the gatekeeper T607I mutation; however, in vitro data suggest that this mutant is sensitive to ponatinib treatment (Garner et al., 2014). Prioritized KIT resistance hotspot residues and their corresponding reported clinical
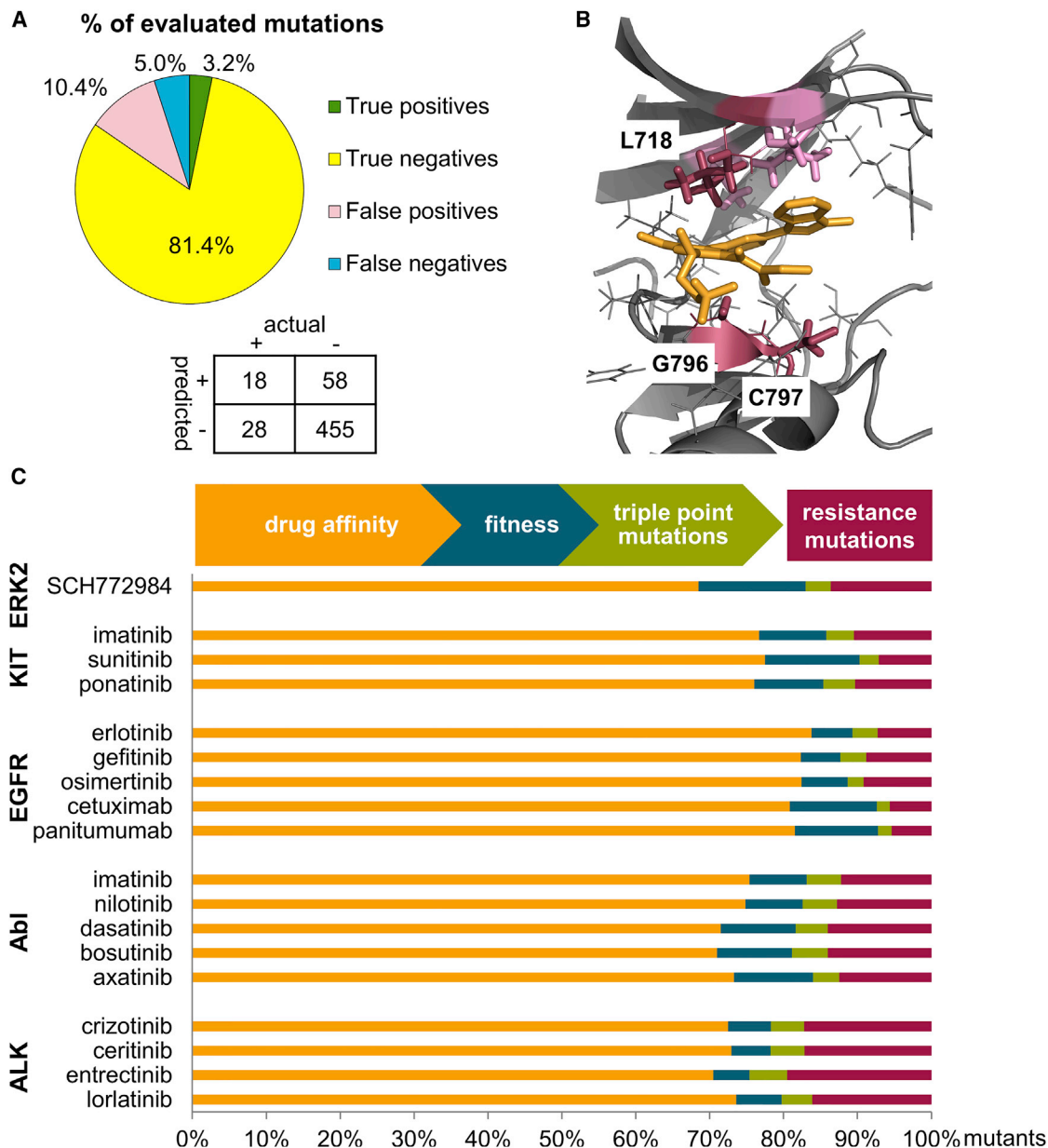
Figure 2. Predicted Resistance Mutations

(A) Performance of the workflow on ERK2-SCH772984. The confusion matrix shows the absolute number of mutations (+, resistant; −, sensitive).

(B) Predicted resistance hotspots that are consistent with clinically observed resistant mutants for the representative case study EGFR (gray) and osimertinib (orange), PDB: 4ZAU (Yosaatmadja et al., 2015). Residues within 5 Å of the ligand are depicted in gray, predicted and clinically observed resistance hotspots are highlighted as crimson sticks and labeled, predicted hotspot residues that have not yet been observed in the clinic are shown as pink sticks; figure created with PyMOL (PyMOL Molecular Graphics System, Version 1.7, Schrödinger, LLC).

(C) Contribution of filtering steps to the identification of resistance mutations. The majority of mutants were discarded because they did not decrease drug affinity in comparison with binding of the endogenous ligand (orange). Mutations were further removed because of abrogated clonal fitness (blue) or because they required triple codon changes to be formed (green). The remaining pool of mutations (crimson) is predicted likely to confer resistance to drug treatment.

resistance mutations are provided in Table 2. A full list of predicted resistance mutations and their relP values is available in Table S3.

EGFR

Applying our workflow to the first-generation EGFR inhibitor erlotinib revealed that G796 is the resistance hotspot with the most predicted mutations. Consistent with this prediction, the G796R mutation with the second-ranked relP, has been reported to weaken the affinity for erlotinib in comparison with the WT enzyme (Avizienyte et al., 2008). T790 is the second-ranked resistance hotspot with gatekeeper mutation T790M (Pao et al., 2005) having the second highest relP. L718, L788, and T854 share the third-ranked hotspot, with three predicted

Table 1. Prediction of ERK2 Mutations

Compound	No. of Pred Mut (No. of All Mut)[a]	Rank 1 Resistance Hotspot (No. of Mut)[b]	Experimentally Confirmed Resistance Mutants at Hotspot Residue[c]	Rank 2 Resistance Hotspot (No. of Mut)[b]	Experimentally Confirmed Resistance Mutants at Hotspot Residue[c]	Rank 3 Resistance Hotspot (No. of Mut)[b]	Experimentally Confirmed Resistance Mutants at Hotspot Residue[c]
SCH772984	76 (559)	Y64 (10)	Y64I[d]	Y36 (9)	Y36R[d]	NA[e]	NA
			Y64L[d]	D111 (9)	Y36N[d]		
			Y64V[d]		Y36Q[d]		
					Y36G[d]		
					Y36I[d]		
					Y36L[d]		
					Y36V[d]		

See also Table S2.

[a]The number of mutants predicted to confer resistance (no. of pred mut) from the initial pool of possible mutants within 5 Å of the ligand (no. of all mut).

[b]Resistance hotspots are identified and ranked according to the number of viable mutants (no. of mut) predicted for each resistance hotspot.

[c]Experimentally observed resistance mutations are highlighted for each resistance hotspot.

[d]Mutants of resistance hotspots were not further ranked based on their relP in this case because Brenan et al. evaluated the mutants in cell lines and the clinical relevance of the mutants for the different cancer types is not known.

[e]NA, not applicable, tied resistance hotspot at rank 2.

mutations each. L788F has been detected in lung adenocarcinoma patient samples highlighting its clinical relevance (Liu et al., 2013); however, similar to all other third-rank mutations, L788F has not yet been associated with clinical resistance to erlotinib. Gefitinib, another first-generation EGFR inhibitor, elicits a similar predicted resistant mutation profile to erlotinib (Table 3).

The third-generation inhibitor osimertinib binds reversibly to EGFR prior to covalent bond formation (Yosaatmadja et al., 2015). Osimertinib was reported to overcome resistance to T790M and, consistent with this finding, the T790M mutation is not predicted to elicit resistance to osimertinib according to our protocol. Residues G796 (14 mutations), C797 (4 mutations), L718, G719, V726, and A743 (3 mutations each) are the top-ranked hotspot residues predicted to confer resistance. During the preparation of this manuscript, osimertinib resistance mutations affecting C797 (Ou et al., 2017b), G796 (Ou et al., 2017b; Zheng et al., 2017), and L718 (Ou et al., 2017a) (Figure 2B) were reported in clinical studies, building confidence in the prospective utility of our approach.

Further to small-molecule EGFR inhibitors, anti-EGFR monoclonal antibodies cetuximab and panitumumab are approved for the treatment of metastatic colorectal cancer. In contrast to small-molecule ATP-competitive kinase inhibitors, cetuximab and panitumumab interact with the extracellular domain of EGFR.

We evaluated the effect of extracellular domain mutations on antibody affinity and affinity of the endogenous ligands EGF and transforming growth factor α (TGF-α)—interactions that have been characterized by protein structural data. Both EGF and TGF-α are peptidic macromolecules with larger interaction surfaces compared with small molecules; thus evaluation of a higher number of possible mutations was necessary to ensure coverage of the extensive protein-protein interaction interface (1,153 and 1,207 mutations for cetuximab and panitumumab, respectively).

For both cetuximab and panitumumab, resistance mutation G441R (residue G465R in the mature protein sequence including the signaling peptide [mps]) has been detected in the clinic (Braig et al., 2015). G441 is the second-ranked resistance hotspot, with mutation G441R ranked first according to relP for both antibodies (Table 3). Confirmed clinical resistance mutations to cetuximab have also been identified at residues S468 (S468R [S492R in the mps]) (Montagut et al., 2012) and I467 (Arena et al., 2015) (I467M [I491M in the mps]); the fourth- and fifth-ranked resistance hotspots in our workflow (Table S4).

Our top-ranked predicted hotspot residues, S418 and G471 have not yet been associated with clinical resistance to cetuximab or panitumumab; however, seven endogenous ligands can activate EGFR (Schneider and Wolf, 2009) of which only EGF and TGF-α interactions are characterized by protein structural data. Residues S418 and G471, while not predicted to affect the interaction of EGFR with endogenous ligands EGF and TGF-α, may be important for interaction with one or more of the remaining five endogenous ligands such that mutation at these residues reduces clonal fitness and persistence. For example, we predict that several mutations affecting EGFR residues S440 and V417 tolerate TGF-α binding, but abrogate EGF binding, and are therefore filtered out in our workflow. Given the lack of structural data for the majority of endogenous EGFR

Table 2. Prioritized KIT Resistance Mutations

Compound	No. of Pred Mut (No. of All Mut)[a]	Rank 1 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutants (Rank relP)[c]	Rank 2 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutants (Rank relP)	Rank 3 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutants (Rank relP)
Imatinib	68 (648)	T670 (8)	T670I (1) (Antonescu et al., 2005; Tamborini et al., 2004) T670E (6) (Wardelmann et al., 2005)	C809 (7)	NR[d]	V668 (6)	NR
Sunitinib	33 (468)	A814 (6)	NR	G596 (5)	NR	A621 (4) V654 (4)	NR
Ponatinib	64 (630)	C809 (8)	NR	T670 (5) C788 (5)	(T670I)[e]	NA[f]	–

See also Table S3.
[a]Number of mutants that have been predicted to confer resistance (no. of pred mut) from the initial pool of possible mutants within 5 Å of the ligand (no. of all mut).
[b]Resistance hotspots are identified and ranked according to the number of viable mutants (no. of mut) predicted for a residue.
[c]RelP was calculated for all resistance hotspot mutations. Clinically observed resistance mutations and their rank according to relP (rank relP) are highlighted for each resistance hotspot.
[d]NR, not reported—none of the predicted mutations has yet been reported to confer resistance to the drug.
[e]T670I was predicted to confer resistance to ponatinib; however, T670I is reported to be sensitive to ponatinib (Garner et al., 2014).
[f]NA, not applicable, tied resistance hotspot at rank 2. The gatekeeper mutation is underlined.

ligands, and the number of potential mutations, the identification of clinically relevant resistance mutations is encouraging. All prioritized EGFR hotspot residues and corresponding reported clinical resistance mutations are summarized in Table 3. A list of all predicted mutations and relPs is provided in Table S4.

Abl

The hotspot residue with the highest number of predicted mutants for the Abl inhibitor imatinib is A380, for which clinically observed occurrences have yet to be reported. The second-rank predicted hotspot is shared by residues V256, A269, and T315. For V256 and T315, resistance mutations V256L (Bamford et al., 2004; Tiribelli et al., 2013), gatekeeper mutation T315I (Gorre et al., 2001), and T315V (Redaelli et al., 2012) have all been identified in the clinic (Table 4). The contribution of mutational signatures to the landscape of mutations has not yet been reported for chronic myeloid leukemia (CML), and therefore the relP for hotspot mutations predicted for Abl residues could not be calculated.

We obtained similar and consistent predictions for nilotinib, dasatinib, and bosutinib (Table 4). Additional resistance mutations affecting L248 are predicted for dasatinib consistent with *in vitro* data (Redaelli et al., 2012). While these mutations have been observed in the clinic (Redaelli et al., 2012), they have not yet been reported as resistant mutants for patients treated with dasatinib. Similarly, the effect of the predicted V256L resistant mutation on nilotinib activity has not been reported.

Axitinib has been reported to overcome resistance to the T315I gatekeeper mutation of Abl (Pemovska et al., 2015). Although not ranked among the top three resistance hotspots, axitinib is predicted to be resistant to T315I. Crystallographic studies of the T315I mutation in complex with axitinib revealed that this mutation causes large conformational changes of Abl compared with WT protein (Pemovska et al., 2015), which may not be adequately captured by our method. The L248R mutation, reported to confer resistance to axatinib *in vitro* (Pemovska et al., 2015), is prioritized by our workflow; however, clinical resistance data for axitinib targeting Abl is not yet available. A summary of the prioritized Abl hotspot residues and their corresponding reported clinical resistance mutations are provided in Table 4. A full list of predicted resistance mutations is provided in Table S5.

ALK

The hotspot residue with the highest frequency of predicted mutations for the first-generation ALK inhibitor crizotinib is G1269 (13 mutants). The clinically observed resistance mutation G1269A (Doebele et al., 2012; Gainor et al., 2016) and G1269S and G1269C mutations (Zhang et al., 2011), which confer resistance *in vitro*, rank at positions four, seven, and eight, respectively, based upon their relP values. For the second-ranked resistance hotspot G1202, 12 mutants were predicted and the known resistance mutation G1202R (Gainor et al., 2016; Katayama et al., 2012) has the second highest relP. Hotspots I1122, G1201, and D1203 rank third equal. While no G1201 mutations are associated with resistance, the D1203N mutation confers resistance *in vitro* (Gainor et al., 2016; Heuckmann et al., 2011) and has been associated with resistance in patients (Zhang et al., 2016). I1122V has been detected in a resistance screen against the second-generation inhibitor brigatinib, and has also

Table 3. Prioritized EGFR Resistance Mutations

Compound	No. of Pred Mut (No. of All Mut)[a]	Rank 1 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutants (Rank relP)[c]	Rank 2 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutants (Rank relP)	Rank 3 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutants (Rank relP)
Erlotinib	34 (469)	G796 (9)	NR[d]	T790 (5)	T790M (2) (Pao et al., 2005)	L718 (3) L788 (3) T854 (3)	NR
Gefitinib	43 (487)	G796 (14)	G796A (3) (Uramoto et al., 2007)	T790 (6)	T790M (2) (Kobayashi et al., 2005; Pao et al., 2005)	T854 (4)	NR
Osimertinib	38 (415)	G796 (14)	G796S (1) (Ou et al., 2017b) G796D (2) (Zheng et al., 2017) G796R (5) (Ou et al., 2017b)	C797 (4)	C797R (3) (Ou et al., 2017b)	L718 (3) G719 (3) V726 (3) A743 (3)	NR
Cetuximab	65 (1,153)	G471 (15)	NR	G441 (12)	G441R (1) (Braig et al., 2015)	S418 (10)	NR
Panitumumab	65 (1,207)	S418 (14)	NR	G441 (12)	G441R (1) (Braig et al., 2015)	G471 (8)	NR

See also Table S4.

[a]The number of mutants predicted to confer resistance (no. of pred mut) from the initial pool of possible mutants within 5 Å of the ligands (no. of all mut).

[b]Resistance hotspots are identified and ranked according to the number of viable mutants (no. of mut) predicted for a residue.

[c]The relP was calculated for all resistance hotspot mutations. Clinically observed resistance mutations and their rank according to relP (rank relP) are highlighted for each resistance hotspot.

[d]NR, not reported—none of the predicted mutations are reported to confer resistance against the drug.

The gatekeeper mutation is underlined.

been confirmed to confer resistance to crizotinib (Ceccon et al., 2015). We generated similar results for the second-generation inhibitor ceritinib (Table 5), although, in this case, the G1269A mutation was not predicted to cause resistance, consistent with *in vitro* and clinical data (Gainor et al., 2016; Shaw et al., 2015).

Entrectinib and lorlatinib are investigational ALK inhibitors currently in clinical trials. Increased half maximal inhibitory concentration values (decreased binding) in comparison with the WT protein have been demonstrated in *in vitro* experiments for several of the mutants prioritized by our workflow. In particular, G1269A (Ardini et al., 2016) and G1202R (Ardini et al., 2016) for entrectinib and G1269A (Gainor et al., 2016; Johnson et al., 2014; Shaw et al., 2015) and D1203N (Gainor et al., 2016) for lorlatinib. In addition, the G1123S and G1123D mutants were identified in an *in vitro* resistance screen for the ALK inhibitor TAE684 (Heuckmann et al., 2011) and in a patient resistant to ceritinib (Toyokawa et al., 2015). However, no data are currently available on entrectinib. The prioritized hotspot residues for ALK and corresponding reported clinical resistance mutations are provided in Table 5. A comprehensive list of all predicted resistance mutations is provided in Table S6.

DISCUSSION

We present a computational workflow to identify clinically relevant drug resistance mutations to targeted cancer therapies, both small molecule and biological. The workflow consists of consecutive filtering steps addressing three factors, which determine whether a mutation is likely to confer resistance in the clinic. Of these, the activity cutoff for protein-drug affinity compared with affinity for the endogenous ligand proved the most important and excluded 75.8% ± 1.1% (n = 18 [all 18 target-interaction partner case studies] ±SEM) of all potential mutants for each drug-target combination (Figure 2C). This filtering step contains a fitness component; all mutants with predicted decreased affinity for the endogenous ligand (i.e., less fit clones), but with equivalent or increased affinity for the drug (i.e., more sensitive to drug treatment) are not progressed. In further considering the potential for disrupted binding of an endogenous ligand to mutated proteins, 8.6% ± 0.7% (n = 18, ±SEM) of mutants are predicted to completely abrogate binding of the endogenous ligand and were discarded. Removing triple-point mutations excluded a further 3.7% ± 0.2% (n = 18, ±SEM) of possible mutants. Interestingly, the majority of remaining mutations predicted to cause resistance (68.1% ± 1.7% [n = 18, ±SEM]) arise from DPM consistent with a similar proportion of DPMs observed in the Brenan dataset (Brenan et al., 2016). While DPMs commonly exhibit a low relP, they may become relevant when the predominant resistance clone harboring an SPM has been eradicated by cancer therapy. Taken together, 11.8% ± 1.0% (n = 18, ±SEM) of all evaluated mutations were considered to cause resistance (Figure 2C).

The pool of potential resistant mutations was further analyzed to identify resistance hotspots where multiple different viable mutants are predicted to cause resistance to the drug under study. Mutants at these hotspots were prioritized based on their relP, which quantifies the relative probability of each specific amino acid mutation in the context of a defined cancer type. Applying this workflow, we correctly classify 84.5% of mutations

Table 4. Prioritized Abl Resistance Mutations

Compound	No. of Pred Mut (No. of All Mut)[a]	Rank 1 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutations[c]	Rank 2 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutations	Rank 3 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutations
Imatinib	66 (540)	A380 (8)	NR[d]	V256 (7) A269 (7) T315 (7)	V256L (Bamford et al., 2004; Tiribelli et al., 2013) T315I (Gorre et al., 2001)	NA[e]	–
Nilotinib	69 (540)	V256 (8) A380 (8)	NR	NA	–	Y253 (6) A269 (6) T315 (6) G321 (6)	T315I (Bamford et al., 2004; Redaelli et al., 2012; Weisberg et al., 2005)
Dasatinib	58 (414)	A380 (12)	NR	L248 (7) A269 (7)	NR	NA	–
Bosutinib	58 (414)	V299 (7) T315 (7) G321 (7)	V299L (Jabbour et al., 2012; Redaelli et al., 2012) T315I (Cortes et al., 2011; Redaelli et al., 2012)	NA	–	NA	–
Axitinib	52 (396)	V256 (8)	NR	G321 (7)	NR	L248 (6) A269 (6) A380 (6)	NR

See also Table S5.

[a]The number of mutants predicted to confer resistance (no. of pred mut) from the initial pool of possible mutants within 5 Å of the ligands (no. of all mut).

[b]Resistance hotspots are identified and ranked according to the number of viable mutants (no. of mut) predicted for a residue.

[c]Clinically observed resistance mutations are highlighted for each resistance hotspot. The relP could not be calculated as signatures for CML were not available.

[d]NR, not reported—none of the predicted mutations were reported to confer resistance. In the case of axitinib, clinical resistance data on Abl are not yet available.

[e]NA, not applicable, tied resistance hotspot at rank 1 or 2. The gatekeeper mutation is underlined.

Table 5. Prioritized ALK Resistance Mutations

Compound	No. of Pred Mut (No. of All Mut)[a]	Rank 1 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutants (Rank relP)[c]	Rank 2 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutants (Rank relP)	Rank 3 Resistance Hotspot (No. of Mut)[b]	Confirmed Clinical Resistance Mutants (Rank relP)[b]
Crizotinib	65 (378)	G1269 (13)	G1269A (4) (Doebele et al., 2012; Gainor et al., 2016)	G1202 (12)	G1202R (2) (Gainor et al., 2016)	I1122 (7) G1201 (7) D1203 (7)	NR[d]
Ceritinib	71 (414)	G1202 (12)	G1202R (2) (Gainor et al., 2016)	G1269 (10) D1203 (10)	D1203N (1) (Gainor et al., 2016)	NA[e]	–
Entrectinib	88 (451)	G1123 (15)	NR	G1269 (13)	NR	G1202 (12)	NR
Lorlatinib	58 (360)	G1269 (13)	NR	G1123 (8) D1203 (8)	NR	NA	–

See also Table S6.

[a]Number of mutants that have been predicted to confer resistance (no. of pred mut) from the initial pool of possible mutants within 5 Å of the ligand (no. of all mut).

[b]Resistance hotspots are identified and ranked according to the number of viable mutants (no. of mut) predicted for each residue.

[c]RelP was calculated for all resistance hotspot mutations. Clinically observed resistance mutations and their rank according to relP (rank relP) are highlighted for each resistance hotspot.

[d]NR, not reported—none of the predicted mutations were reported to confer resistance against the drug. In the case of entrectinib and lorlatinib clinical resistance data are not yet available.

[e]NA, not applicable, tied resistance hotspot at rank 1 or 2.

in a comprehensive ERK2 mutation dataset and identify almost 40% of mutants conferring resistance to the ERK2 inhibitor SCH772984. Furthermore, we identify clinically observed drug resistance mutations within the top three predicted hotspot residues for first-generation compounds in all cases studied and for most second- and third-generation compounds where clinically observed resistance mutations have been reported. Importantly, in all cases studied, except ALK, the gatekeeper mutation is predicted among the top three resistance hotspots. In the ALK case study, two mutations that are also commonly observed in the clinic are highly ranked by our method. For the osimertinib-EGFR drug-target pair, we predicted resistance mutations before they were confirmed by clinical reports.

Our workflow highlights mutations not yet observed in the clinic, or that may constitute FP predictions. The number of mutants progressed to further evaluation can be user defined. In the cases exemplified here, we prioritize three mutations with the highest relP for the top three resistance hotspots and significantly narrow the pool of potential resistance mutations for each drug from ~350–1,200 possibilities to 9. While it may not be practical to further explore many hundreds of possible mutations, we suggest that a set of 9 prioritized mutants is more amenable to experimental testing within a drug discovery project or clinical setting. Thus, this workflow facilitates more focused monitoring of potential resistant mutations, as well as the design of next-generation compounds sensitive to likely resistant mutants.

Notably, our workflow evaluates the effect of a mutation on drug binding with respect to the WT protein. In some cases, exquisite drug affinity for the WT protein may mitigate a mutation-driven loss of potency. For example, our workflow correctly predicted the reduced affinity of lorlatinib for ALK G1269A; however, the exquisite potency of lorlatinib for WT ALK may render the relative loss of affinity for ALK mutants inconsequential (Gainor et al., 2016; Johnson et al., 2014; Shaw et al., 2015).

The workflow is dependent upon the availability and quality of both protein-ligand structural information and mutational signatures across diverse tumor types. For example, it is not clear which mutational signatures operate in CML, and, as a consequence, we could not determine the relP for mutants predicted to interfere with binding of Abl inhibitors in CML.

The approach presented here investigates the effects of protein mutation on drug binding mode as characterized by the input protein-ligand structure. How such mutations affect the overall conformation of the target protein and/or drug-target complex is not encompassed by our method. Extension of the approach by considering the effect of residue mutation on global protein conformation and by inclusion of structural models for highly homologous protein families would further expand the potential impact and is the aim of our future studies. For example, FP mutations that stabilize inactive protein conformations that are unlikely to persist could be excluded, and mutations that further stabilize an active protein conformation could be included. Furthermore, drug-resistant mutations distant from the binding site, including those which influence protein flexibility and conformation, which are beyond the scope of our current method, could also be evaluated.

This workflow includes three critical determinants of clinically relevant drug-resistant mutations. Importantly, the workflow can

be further adapted to test more specific hypotheses. For example, Osprey parameters can be changed to include larger backbone movements and/or to include multiple ligand rotamers; furthermore, the number of prioritized mutants can also be user-defined.

In conclusion, we have developed and validated a computational method to prospectively identify clinically relevant drug-resistance mutations. We suggest that this approach can have a significant impact on the design and development of targeted therapies by proactively signposting drug resistance hotspots. Prior knowledge of resistant mutants enables their timely detection in patients and the early development of effective treatment options against the resistant tumor cell population.

SIGNIFICANCE

Although molecularly targeted cancer therapies have shown great success in the clinic, drug resistance has emerged as the major challenge. Resistance mutations are commonly identified and characterized during clinical evaluation, often resulting in a reactive approach to tackling drug resistance. We report a computational method to prospectively identify drug resistance mutations during the design phase of potential therapeutics. This approach enables early signposting of likely resistance hotspots and supports more focused monitoring of potential emergent resistant clones as well as the timely development of alternative treatment options.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- ● KEY RESOURCES TABLE
- ● CONTACT FOR REAGENT AND RESOURCE SHARING
- ● DATA AND SOFTWARE AVAILABILITY
- ● METHOD DETAILS
  - ○ Structure Selection and Preparation
  - ○ Evaluation of Ligand Affinity
  - ○ Calculation of relP
- ● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and eight tables and can be found with this article online at https://doi.org/10.1016/j.chembiol.2018.07.013.

AUTHOR CONTRIBUTIONS

T.K. conducted the experiments. J.B. and T.K. designed the experiments and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Garrett et al., 2002; Stamos et al., 2002; Mol et al., 2003; Mol et al., 2004; Li et al., 2005; Levinson et al., 2006; Tokarski et al., 2006; Yun et al., 2007; Cowan-Jacob et al., 2007; Gajiwala et al., 2009; Lee et al., 2010; Lu et al., 2010; Cui et al., 2011; Garai et al., 2012; Levinson and Boxer, 2012; Chaikuad et al., 2014; Friboulet et al., 2014; Sander et al., 2015; Menichincheri et al., 2016; Sickmier et al., 2016.

REFERENCES

Abbosh, C., Birkbak, N.J., Wilson, G.A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D.A., Veeriah, S., Rosenthal, R., et al. (2017). Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature *545*, 446–451.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Borresen-Dale, A.-L., et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Antonescu, C.R., Besmer, P., Guo, T., Arkun, K., Hom, G., Koryotowski, B., Leversha, M.A., Jeffrey, P.D., Desantis, D., Singer, S., et al. (2005). Acquired resistance to imatinib in gastrointestinal stromal tumor occurs through secondary gene mutation. Clin. Cancer Res. *11*, 4182–4190.

Ardini, E., Menichincheri, M., Banfi, P., Bosotti, R., De Ponti, C., Pulci, R., Ballinari, D., Ciomei, M., Texido, G., Degrassi, A., et al. (2016). Entrectinib, a pan-TRK, ROS1, and ALK inhibitor with activity in multiple molecularly defined cancer indications. Mol. Cancer Ther. *15*, 628–639.

Arena, S., Bellosillo, B., Siravegna, G., Martínez, A., Cañadas, I., Lazzari, L., Ferruz, N., Russo, M., Misale, S., González, I., et al. (2015). Emergence of multiple EGFR extracellular mutations during cetuximab treatment in colorectal cancer. Clin. Cancer Res. *21*, 2157–2166.

Avizienyte, E., Ward, Richard A., and Garner, Andrew P. (2008). Comparison of the EGFR resistance mutation profiles generated by EGFR-targeted tyrosine kinase inhibitors and the impact of drug combinations. Biochem. J. *415*, 197–206.

Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P.A., Stratton, M.R., and Wooster, R. (2004). The COSMIC (catalogue of somatic mutations in cancer) database and website. Br. J. Cancer *91*, 355–358.

Barouch-Bentov, R., and Sauer, K. (2011). Mechanisms of drug resistance in kinases. Expert Opin. Invest. Drugs *20*, 153–208.

Bouaoun, L., Sonkin, D., Ardin, M., Hollstein, M., Byrnes, G., Zavadil, J., and Olivier, M. (2016). TP53 variations in human cancers: new lessons from the IARC TP53 database and genomics data. Hum. Mutat. *37*, 865–876.

Braig, F., März, M., Schieferdecker, A., Schulte, A., Voigt, M., Stein, A., Grob, T., Alawi, M., Indenbirken, D., Kriegs, M., et al. (2015). Epidermal growth factor receptor mutation mediates cross-resistance to panitumumab and cetuximab in gastrointestinal cancer. Oncotarget *6*, 12035–12047.

Brenan, L., Andreev, A., Cohen, O., Pantel, S., Kamburov, A., Cacchiarelli, D., Persky, Nicole S., Zhu, C., Bagul, M., Goetz, E.M., et al. (2016). Phenotypic characterization of a comprehensive set of MAPK1/ERK2 missense mutants. Cell Rep. *17*, 1171–1183.

Case, D.A., Betz, R.M., Cerutti, D.S., Cheatham, T.E., III, Darden, T.A., Duke, R.E., Giese, T.J., Gohlke, H., Goetz, A.W., Homeyer, N., et al. (2016). AMBER 2016 (University of California, San Francisco).

Ceccon, M., Mologni, L., Giudici, G., Piazza, R., Pirola, A., Fontana, D., and Gambacorti-Passerini, C. (2015). Treatment efficacy and resistance mechanisms using the second-generation ALK inhibitor AP26113 in human NPM-ALK-positive anaplastic large cell lymphoma. Mol. Cancer Res. *13*, 775–783.

Chaikuad, A., Tacconi, E.M., Zimmer, J., Liang, Y., Gray, N.S., Tarsounas, M., and Knapp, S. (2014). A unique inhibitor binding site in ERK1/2 is associated with slow binding kinetics. Nat. Chem. Biol. *10*, 853–860.

Chen, C.-Y., Georgiev, I., Anderson, A.C., and Donald, B.R. (2009). Computational structure-based redesign of enzyme activity. Proc. Natl. Acad. Sci. USA *106*, 3764–3769.

Cortes, J.E., Kantarjian, H.M., Brümmendorf, T.H., Kim, D.-W., Turkina, A.G., Shen, Z.-X., Pasquini, R., Khoury, H.J., Arkin, S., Volkert, A., et al. (2011). Safety and efficacy of bosutinib (SKI-606) in chronic phase Philadelphia chromosome-positive chronic myeloid leukemia patients with resistance or intolerance to imatinib. Blood *118*, 4567–4576.

Cowan-Jacob, S.W., Fendrich, G., Floersheimer, A., Furet, P., Liebetanz, J., Rummel, G., Rheinberger, P., Centeleghe, M., Fabbro, D., and Manley, P.W. (2007). Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. Acta Crystallogr. D Biol. Crystallogr. *63*, 80–93.

Cui, J.J., Tran-Dubé, M., Shen, H., Nambu, M., Kung, P.-P., Pairish, M., Jia, L., Meng, J., Funk, L., Botrous, I., et al. (2011). Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal–epithelial transition factor (c-MET) kinase and anaplastic lymphoma kinase (ALK). J. Med. Chem. *54*, 6342–6363.

Doebele, R.C., Pilling, A.B., Aisner, D.L., Kutateladze, T.G., Le, A.T., Weickhardt, A.J., Kondo, K.L., Linderman, D.J., Heasley, L.E., Franklin, W.A., et al. (2012). Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. Clin. Cancer Res. *18*, 1472–1482.

Fowler, P.W., Cole, K., Gordon, N.C., Kearns, A.M., Llewelyn, M.J., Peto, T.E.A., Crook, D.W., and Walker, A.S. (2018). Robust prediction of resistance to trimethoprim in *Staphylococcus aureus*. Cell Chem. Biol. *25*, 339–349.

Frey, K.M., Georgiev, I., Donald, B.R., and Anderson, A.C. (2010). Predicting resistance mutations using protein design algorithms. Proc. Natl. Acad. Sci. USA *107*, 13707–13712.

Friboulet, L., Li, N., Katayama, R., Lee, C.C., Gainor, J.F., Crystal, A.S., Michellys, P.-Y., Awad, M.M., Yanagitani, N., Kim, S., et al. (2014). The ALK inhibitor ceritinib overcomes crizotinib resistance in non-small cell lung cancer. Cancer Discov. *4*, 662–673.

Gainor, J.F., Dardaei, L., Yoda, S., Friboulet, L., Leshchiner, I., Katayama, R., Dagogo-Jack, I., Gadgeel, S., Schultz, K., Singh, M., et al. (2016). Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK-rearranged lung cancer. Cancer Discov. *6*, 1118–1133.

Gainza, P., Roberts, K.E., Georgiev, I., Lilien, R.H., Keedy, D.A., Chen, C.-Y., Reza, F., Anderson, A.C., Richardson, D.C., Richardson, J.S., and Donald, B.R. (2013). Osprey: protein design with ensembles, flexibility, and provable algorithms. In Methods in Enzymology, A.E. Keating, ed. (Academic Press), pp. 87–107.

Gajiwala, K.S., Wu, J.C., Christensen, J., Deshmukh, G.D., Diehl, W., DiNitto, J.P., English, J.M., Greig, M.J., He, Y.-A., Jacques, S.L., et al. (2009). KIT kinase mutants show unique mechanisms of drug resistance to imatinib and sunitinib in gastrointestinal stromal tumor patients. Proc. Natl. Acad. Sci. USA *106*, 1542–1547.

Garai, Á., Zeke, A., Gógl, G., Törő, I., Fördős, F., Blankenburg, H., Bárkai, T., Varga, J., Alexa, A., Emig, D., et al. (2012). Specificity of linear motifs that bind to a common mitogen-activated protein kinase docking groove. Sci. Signal. *5*, ra74.

Garner, A.P., Gozgit, J.M., Anjum, R., Vodala, S., Schrock, A., Zhou, T., Serrano, C., Eilers, G., Zhu, M., Ketzer, J., et al. (2014). Ponatinib inhibits polyclonal drug-resistant KIT oncoproteins and shows therapeutic potential in heavily pretreated gastrointestinal stromal tumor (GIST) patients. Clin. Cancer Res. *20*, 5745–5755.

Garrett, T.P., McKern, N.M., Lou, M., Elleman, T.C., Adams, T.E., Lovrecz, G.O., Zhu, H.-J., Walker, F., Frenkel, M.J., Hoyne, P.A., et al. (2002). Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α. Cell *110*, 763–773.

Gil, J., and Rodriguez, T. (2016). Cancer: the transforming power of cell competition. Curr. Biol. *26*, R164–R166.

Gorre, M.E., Mohammed, M., Ellwood, K., Hsu, N., Paquette, R., Rao, P.N., and Sawyers, C.L. (2001). Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science *293*, 876–880.

Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. Nature *481*, 306–313.

Gurden, M.D., Westwood, I.M., Faisal, A., Naud, S., Cheung, K.-M.J., McAndrew, C., Wood, A., Schmitt, J., Boxall, K., Mak, G., et al. (2015). Naturally occurring mutations in the MPS1 gene predispose cells to kinase inhibitor drug resistance. Cancer Res. *75*, 3340–3354.

Heuckmann, J.M., Hölzel, M., Sos, M.L., Heynck, S., Balke-Want, H., Koker, M., Peifer, M., Weiss, J., Lovly, C.M., Grütter, C., et al. (2011). ALK mutations conferring differential resistance to structurally diverse ALK inhibitors. Clin. Cancer Res. *17*, 7394–7401.

Holohan, C., Van Schaeybroeck, S., Longley, D.B., and Johnston, P.G. (2013). Cancer drug resistance: an evolving paradigm. Nat. Rev. Cancer *13*, 714–726.

Hosseini, A., Alibés, A., Noguera-Julian, M., Gil, V., Paredes, R., Soliva, R., Orozco, M., and Guallar, V. (2016). Computational prediction of HIV-1 resistance to protease inhibitors. J. Chem. Inf. Model. *56*, 915–923.

Inukai, M., Toyooka, S., Ito, S., Asano, H., Ichihara, S., Soh, J., Suehisa, H., Ouchida, M., Aoe, K., Aoe, M., et al. (2006). Presence of epidermal growth factor receptor gene T790M mutation as a minor clone in non-small cell lung cancer. Cancer Res. *66*, 7854–7858.

Jabbour, E., Morris, V., Kantarjian, H., Yin, C.C., Burton, E., and Cortes, J. (2012). Characteristics and outcomes of patients with V299L BCR-ABL kinase domain mutation after therapy with tyrosine kinase inhibitors. Blood *120*, 3382–3383.

Jamal-Hanjani, M., Wilson, G.A., McGranahan, N., Birkbak, N.J., Watkins, T.B.K., Veeriah, S., Shafi, S., Johnson, D.H., Mitter, R., Rosenthal, R., et al. (2017). Tracking the evolution of non-small-cell lung cancer. N. Engl. J. Med. *376*, 2109–2121.

Johnson, T.W., Richardson, P.F., Bailey, S., Brooun, A., Burke, B.J., Collins, M.R., Cui, J.J., Deal, J.G., Deng, Y.-L., Dinh, D., et al. (2014). Discovery of (10R)-7-amino-12-fluoro-2,10,16-trimethyl-15-oxo-10,15,16,17-tetrahydro-2H-8,4-(metheno)pyrazolo[4,3-h][2,5,11]-benzoxadiazacyclotetradecine-3-carbonitrile (PF-06463922), a macrocyclic inhibitor of anaplastic lymphoma kinase (ALK) and c-ros oncogene 1 (ROS1) with preclinical brain exposure and broad-spectrum potency against ALK-resistant mutations. J. Med. Chem. *57*, 4720–4744.

Kamasani, S., Akula, S., Sivan, S.K., Manga, V., Duyster, J., Vudem, D.R., and Kancha, R.K. (2017). Computational analysis of ABL kinase mutations allows predicting drug sensitivity against selective kinase inhibitors. Tumor Biol. *39*, 1–7.

Katayama, R., Shaw, A.T., Khan, T.M., Mino-Kenudson, M., Solomon, B.J., Halmos, B., Jessop, N.A., Wain, J.C., Yeo, A.T., Benes, C., et al. (2012). Mechanisms of acquired crizotinib resistance in ALK-rearranged lung cancers. Sci. Transl. Med. *4*, 1–12.

Kobayashi, S., Boggon, T.J., Dayaram, T., Jänne, P.A., Kocher, O., Meyerson, M., Johnson, B.E., Eck, M.J., Tenen, D.G., and Halmos, B. (2005). EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N. Engl. J. Med. *352*, 786–792.

Lee, C.C., Jia, Y., Li, N., Sun, X., Ng, K., Ambing, E., Gao, M.-Y., Hua, S., Chen, C., Kim, S., et al. (2010). Crystal structure of the ALK (anaplastic lymphoma kinase) catalytic domain. Biochem. J. *430*, 425–437.

Levinson, N.M., and Boxer, S.G. (2012). Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain. PLoS One *7*, 1–10.

Levinson, N.M., Kuchment, O., Shen, K., Young, M.A., Koldobskiy, M., Karplus, M., Cole, P.A., and Kuriyan, J. (2006). A Src-like inactive conformation in the Abl tyrosine kinase domain. PLoS Biol. *4*, 0753–0767.

Li, S., Schmitz, K.R., Jeffrey, P.D., Wiltzius, J.J., Kussie, P., and Ferguson, K.M. (2005). Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. Cancer Cell *7*, 301–311.

Liu, Y., Wu, B.-Q., Zhong, H.-H., Hui, P., and Fang, W.-G. (2013). Screening for EGFR and KRAS mutations in non-small cell lung carcinomas using DNA extraction by hydrothermal pressure coupled with PCR-based direct sequencing. Int. J. Clin. Exp. Pathol. 6, 1880–1889.

Lorenzo, F., Nishii, K., Monma, F., Kuwagata, S., Usui, E., and Shiku, H. (2006). Mutational analysis of the KIT gene in myelodysplastic syndrome (MDS) and MDS-derived leukemia. Leuk. Res. 30, 1235–1239.

Lu, C., Mi, L.-Z., Grey, M.J., Zhu, J., Graef, E., Yokoyama, S., and Springer, T.A. (2010). Structural evidence for loose linkage between ligand binding and kinase activation in the epidermal growth factor receptor. Mol. Cell. Biol. 30, 5432–5443.

Martínez-Jiménez, F., Overington, J.P., Al-Lazikani, B., and Marti-Renom, M.A. (2017). Rational design of non-resistant targeted cancer therapies. Sci. Rep. 7, 1–18.

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.

Mol, C.D., Lim, K.B., Sridhar, V., Zou, H., Chien, E.Y., Sang, B.-C., Nowakowski, J., Kassel, D.B., Cronin, C.N., and McRee, D.E. (2003). Structure of a c-Kit product complex reveals the basis for kinase transactivation. J. Biol. Chem. 278, 31461–31464.

Mol, C.D., Dougan, D.R., Schneider, T.R., Skene, R.J., Kraus, M.L., Scheibe, D.N., Snell, G.P., Zou, H., Sang, B.-C., and Wilson, K.P. (2004). Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. J. Biol. Chem. 279, 31655–31663.

Molecular Operating Environment (MOE), 2015.1001; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2017.

Montagut, C., Dalmases, A., Bellosillo, B., Crespo, M., Pairet, S., Iglesias, M., Salido, M., Gallen, M., Marsters, S., Tsai, S.P., et al. (2012). Identification of a mutation in the extracellular domain of the epidermal growth factor receptor conferring cetuximab resistance in colorectal cancer. Nat. Med. 18, 221–223.

Ou, Q., Wu, X., Bao, H., Tong, X., Wang, X., Zhang, X., Chang, Z., Mao, Y., Shao, Y., and Zhou, C. (2017a). Investigating novel resistance mechanisms to third generation EGFR TKI osimertinib in non-small cell lung cancer patients using next generation sequencing. J. Clin. Oncol. 35, 2572.

Ou, S.-H.I., Cui, J., Schrock, A.B., Goldberg, M.E., Zhu, V.W., Albacker, L., Stephens, P.J., Miller, V.A., and Ali, S.M. (2017b). Emergence of novel and dominant acquired EGFR solvent-front mutations at Gly796 (G796S/R) together with C797S/R and L792F/H mutations in one EGFR (L858R/T790M) NSCLC patient who progressed on osimertinib. Lung Cancer 108, 228–231.

Pandurangan, A.P., Ascher, D.B., Thomas, S.E., and Blundell, T.L. (2017). Genomes, structural biology and drug discovery: combating the impacts of mutations in genetic disease and antibiotic resistance. Biochem. Soc.Trans. 45, 303–311.

Pao, W., Miller, V.A., Politi, K.A., Riely, G.J., Somwar, R., Zakowski, M.F., Kris, M.G., and Varmus, H. (2005). Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. PLoS Med. 2, 0225–0235.

Paul, N., Carabet, L.A., Lallous, N., Yamazaki, T., Gleave, M.E., Rennie, P.S., and Cherkasov, A. (2016). Cheminformatics modeling of adverse drug responses by clinically relevant mutants of human androgen receptor. J. Chem. Inf. Model. 56, 2507–2516.

Pemovska, T., Johnson, E., Kontro, M., Repasky, G.A., Chen, J., Wells, P., Cronin, C.N., McTigue, M., Kallioniemi, O., Porkka, K., et al. (2015). Axitinib effectively inhibits BCR-ABL1(T315I) with a distinct binding conformation. Nature 519, 102–105.

Redaelli, S., Mologni, L., Rostagno, R., Piazza, R., Magistroni, V., Ceccon, M., Viltadi, M., Flynn, D., and Gambacorti-Passerini, C. (2012). Three novel patient-derived BCR/ABL mutants show different sensitivity to second and third generation tyrosine kinase inhibitors. Am. J. Hematol. 87, E125–E128.

Reeve, S.M., Gainza, P., Frey, K.M., Georgiev, I., Donald, B.R., and Anderson, A.C. (2015). Protein design algorithms predict viable resistance to an experimental antifolate. Proc. Natl. Acad. Sci. USA 112, 749–754.

Schrödinger Release 2014-2: Maestro, Schrödinger, LLC, New York, NY, 2014.

Roche-Lestienne, C., Soenen-Cornu, V., Grardel-Duflos, N., Laï, J.-L., Philippe, N., Facon, T., Fenaux, P., and Preudhomme, C. (2002). Several types of mutations of the Abl gene can be found in chronic myeloid leukemia patients resistant to STI571, and they can pre-exist to the onset of treatment. Blood 100, 1014–1018.

Sander, T., Freyss, J., von Korff, M., and Rufener, C. (2015). DataWarrior: an open-source program for chemistry aware data visualization and analysis. J. Chem. Inf. Model. 55, 460–473.

Sastry, G.M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J. Comput. Aided Mol. Des. 27, 221–234.

Schneider, M.R., and Wolf, E. (2009). The epidermal growth factor receptor ligands at a glance. J. Cell. Physiol. 218, 460–466.

Shaw, A.T., Friboulet, L., Leshchiner, I., Gainor, J.F., Bergqvist, S., Brooun, A., Burke, B.J., Deng, Y.-L., Liu, W., Dardaei, L., et al. (2015). Resensitization to crizotinib by the lorlatinib ALK resistance mutation L1198F. N. Engl. J. Med. 374, 54–61.

Sickmier, E.A., Kurzeja, R.J., Michelsen, K., Vazir, M., Yang, E., and Tasker, A.S. (2016). The panitumumab EGFR complex reveals a binding mechanism that overcomes cetuximab induced resistance. PLoS One 11, 1–11.

Sprouffske, K., Merlo, L.M., Gerrish, P.J., Maley, C.C., and Sniegowski, P.D. (2012). Cancer in light of experimental evolution. Curr. Biol. 22, R762–R771.

Stamos, J., Sliwkowski, M.X., and Eigenbrot, C. (2002). Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. J. Biol. Chem. 277, 46265–46272.

Tamborini, E., Bonadiman, L., Greco, A., Albertini, V., Negri, T., Gronchi, A., Bertulli, R., Colecchia, M., Casali, P.G., Pierotti, M.A., and Pilotti, S. (2004). A new mutation in the KIT ATP pocket causes acquired resistance to imatinib in a gastrointestinal stromal tumor patient. Gastroenterology 127, 294–299.

Tiribelli, M., Latagliata, R., Luciano, L., Castagnetti, F., Gozzini, A., Cambrin, G.R., Annunziata, M., Stagno, F., Pregno, P., Albano, F., et al. (2013). Impact of BCR-ABL mutations on response to dasatinib after imatinib failure in elderly patients with chronic-phase chronic myeloid leukemia. Ann. Hematol. 92, 179–183.

Tokarski, J.S., Newitt, J.A., Chang, C.Y., Cheng, J.D., Wittekind, M., Kiefer, S.E., Kish, K., Lee, F.Y., Borzillerri, R., Lombardo, L.J., et al. (2006). The structure of dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against imatinib-resistant ABL mutants. Cancer Res. 66, 5790–5797.

Toyokawa, G., Inamasu, E., Shimamatsu, S., Yoshida, T., Nosaki, K., Hirai, F., Yamaguchi, M., Seto, T., Takenoyama, M., and Ichinose, Y. (2015). Identification of a novel ALK G1123S mutation in a patient with ALK-rearranged non-small-cell lung cancer exhibiting resistance to ceritinib. J. Thorac. Oncol. 10, e55–e57.

Uramoto, H., Uchiumi, T., Izumi, H., Kohno, K., Oyama, T., Sugio, K., and Yasumoto, K. (2007). A new mechanism for primary resistance to gefitinib in lung adenocarcinoma: the role of a novel G796A mutation in exon 20 of EGFR. Anticancer Res. 27, 2297–2303.

Wardelmann, E., Thomas, N., Merkelbach-Bruse, S., Pauls, K., Speidel, N., Büttner, R., Bihl, H., Leutner, C.C., Heinicke, T., and Hohenberger, P. (2005). Acquired resistance to imatinib in gastrointestinal stromal tumours caused by multiple KIT mutations. Lancet Oncol. 6, 249–251.

Weisberg, E., Manley, P.W., Breitenstein, W., Brüggen, J., Cowan-Jacob, S.W., Ray, A., Huntly, B., Fabbro, D., Fendrich, G., Hall-Meyers, E., et al. (2005). Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. Cancer Cell 7, 129–141.

Yosaatmadja, Y., Silva, S., Dickson, J.M., Patterson, A.V., Smaill, J.B., Flanagan, J.U., McKeage, M.J., and Squire, C.J. (2015). Binding mode of the breakthrough inhibitor AZD9291 to epidermal growth factor receptor revealed. J. Struct. Biol. *192*, 539–544.

Yun, C.-H., Boggon, T.J., Li, Y., Woo, M.S., Greulich, H., Meyerson, M., and Eck, M.J. (2007). Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. Cancer Cell *11*, 217–227.

Zhang, S., Wang, F., Keats, J., Zhu, X., Ning, Y., Wardwell, S.D., Moran, L., Mohemmad, Q.K., Anjum, R., Wang, Y., et al. (2011). Crizotinib-resistant mutants of EML4-ALK identified through an accelerated mutagenesis screen. Chem. Biol. Drug Des. *78*, 999–1005.

Zhang, S., Anjum, R., Squillace, R., Nadworny, S., Zhou, T., Keats, J., Ning, Y., Wardwell, S.D., 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.

Zheng, D., Hu, M., Bai, Y., Zhu, X., Lu, X., Wu, C., Wang, J., Liu, L., Wang, Z., Ni, J., et al. (2017). EGFR G796D mutation mediates resistance to osimertinib. Oncotarget *8*, 49671–49679.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Crystal structures of target-ligand complexes	The Protein Data Bank	Table S7; http://www.rcsb.org/pdb/home/home.do
Mutation signatures	Alexandrov et al., 2013; This study	ftp://ftp.sanger.ac.uk/pub/cancer/AlexandrovEtAl/ Table S8
Coding sequence for wt targets	COSMIC, Bamford et al., 2004	http://cancer.sanger.ac.uk/cosmic
IARC P53 Database R18	Bouaoun et al., 2016	http://p53.iarc.fr/
TRACERx	Jamal-Hanjani et al., 2017; Abbosh et al., 2017	
COSMIC Resistance Mutations v83	COSMIC, Bamford et al., 2004	http://cancer.sanger.ac.uk/cosmic
COSMIC Mutation Data v83	Bamford et al., 2004	http://cancer.sanger.ac.uk/cosmic
Software and Algorithms		
Maestro version 9.8.016	Schrödinger	https://www.schrodinger.com/maestro
MOE 2015.1001	Chemical Computing Group	https://www.chemcomp.com/
Osprey version 2.2beta	Chen et al., 2009; Gainza et al., 2013	http://www.cs.duke.edu/donaldlab/osprey.php
AmberTools16	Case et al., 2016	http://ambermd.org/AmberTools16-get.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Julian Blagg (Julian.blagg@icr.ac.uk).

DATA AND SOFTWARE AVAILABILITY

Data are available upon request to the Lead Contact.

METHOD DETAILS

Structure Selection and Preparation

When available, crystal structures of wild type protein were used. Formation of the Lys-Glu salt bridge was the minimum requirement for kinase co-factor structures to be considered representative of an active conformation. Due to the paucity of high resolution ATP-bound kinase protein structures, we employed ADP or ATP analogue complexes for all kinase structures representative of endogenous co-factor binding. A list of employed PDB entries is provided in Table S7. All crystal structures were prepared using the Preparation wizard (Sastry et al., 2013) in Maestro version 9.8.016 (Schrödinger Release, 2014). Bond orders were assigned and hydrogens were added. Zero-order bonds to metals and disulphide bonds were created and selenomethionines were converted to methionines. Missing side chains were added and all waters beyond 5 Å from heteroatoms, buffer compounds, and additional chains were deleted. Ionization states were generated using Epik. H-bonds were assigned, water conformations were sampled and structures minimized using default settings.

Residues within 12 Å of the ligand were selected using the selector tool in MOE 2015.1001 (Molecular Operating Environment (MOE), 2017) and only selected residues were included in the input structure. As described in the Osprey documentation, the protonation state of histidines was defined in the pdb input file, all HETATM identifiers were replaced by ATOM, and chain information was deleted.

Evaluation of Ligand Affinity

Osprey version 2.2beta (Chen et al., 2009; Gainza et al., 2013), using the Amber94 force field, was employed to evaluate the impact of protein mutation on ligand binding. All non-peptide interaction partners were parameterized using Antechamber (AmberTools16 (Case et al., 2016)) as described in the Osprey documentation and only the selected crystal structure protein conformation was used. Small molecule ligands were allowed to rotate and translate during the calculations.

Residues within 5 Å of the ligand were systematically mutated to all other possible amino acids, except if they were known to be crucial for catalytic activity; namely the AspPheGly motif, catalytic Asp, and Lys-Glu salt bridge in kinases. All three His protonation

states were included; His was only considered as a resistant mutation when all protonation states were predicted to confer resistance. Mutation to Pro is not supported by the Osprey package. In cases where ligands form specific water-mediated interactions with the protein, the water molecules were included and defined as co-factor as suggested in the Osprey documentation. Only one residue at a time was allowed to mutate and WT rotamers were included for all mutable residues. Residue positions were investigated either individually or in pairs; this led to multiple WT K* scores per protein-ligand complex and the average WT (avwt) score was used henceforth. The ratio of the endogenous ligand to drug K* score (= K* ratio) was calculated (Frey et al., 2010; Reeve et al., 2015) to evaluate the effect of each possible mutation on drug binding in relation to binding of the endogenous ligand. The log score of the K* ratio was subsequently used. The range of WT scores ($range_{wt}$) across log-units was used to determine the variation range of the method for each input structure which results from the quality of the input model. (Figure S1). A decrease in drug affinity for a mutant compared to the WT protein has the potential to cause resistance, the activity cut-off for mutants was defined as a log K* ratio higher than the average WT protein value plus the inherent variation range of the method as described in Equation 1:

$$cut-off = \log(K^*ratio_{avwt}) + range_{wt} \qquad \text{(Equation 1)}$$

Every mutation retrieving a value higher than the cut-off was considered as conferring resistance.

The avwt score was used to define K* score values for residue positions that were only investigated for either endogenous ligand or drug. This was the case if a residue was within 5 Å of one but not the other.

Calculation of relP

The mutation type probabilities were taken from Alexandrov et al. (Alexandrov et al., 2013). The relevant signatures were extracted based on the cancer type for which the drug was approved. In detail, the signatures operating in stomach cancer and lung adenocarcinoma were used to calculate relative probabilities for KIT/ALK and EGFR small molecule drugs, respectively. Colorectal cancer signatures were employed for EGFR-antibody complexes. No signatures for CML (relevant for inhibitors of Abl) have been reported.

The original values of the signatures (x_s) were normalized according to their total contribution to mutational load and the number of samples in which the signature could be detected (c_s) (Figure S2). The normalized signature probabilities of all relevant signatures were added to give the overall probability for a specific base exchange in a particular cancer type (=SPM probability, pSPM) as described in Equation 2:

$$pSPM = \sum_{s_1}^{s_n} c_s * x_s \qquad \text{(Equation 2)}$$

where:

s_1 and s_n are the first and n^{th} signature contributing to the mutational load in a particular cancer type.

The dataset of cancer-specific SPM probabilities is provided in Table S8 in the Supplemental Information.

To calculate the relP for a specific amino acid missense mutation, the coding sequence for each amino acid residue and their 5' and 3' neighbouring bases were extracted from the COSMIC database (Bamford et al., 2004). Signature-derived pSPMs (as described above) were used to define the relP for a specific amino acid exchange according to Equation 3:

$$relP = \sum_{t_1}^{t_n} pSPM_1 * pSPM_2 \qquad \text{(Equation 3)}$$

where:

t_1 and t_n are the first and n^{th} combination of base exchanges resulting in the desired protein residue change.

$pSPM_1$ and $pSPM_2$ are the overall probability for a specific base exchange 1 and 2 in a particular cancer type. DPMs have been reported in cancer patients with similar frequency to insertions (Table S1) and were therefore included in the workflow. Please note, that only $pSPM_1$ is required if a protein residue mutation can be facilitated by a single base exchange (SPM), whereas both $pSPM_1$ and $pSPM_2$ are used for a DPM.

Whenever multiple different mutations led to the same amino acid exchange, the individual SPM probabilities were added. The calculation of relP, exemplified on the EGFR T790M mutation in adenocarcinoma, is depicted in Figure S2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Microsoft Excel was used to calculate the mean ± SEM across all 18 target-interaction partner case studies (=n) as reported in the Discussion.

Supplemental Information

Combining Mutational Signatures, Clonal Fitness,

and Drug Affinity to Define Drug-Specific

Resistance Mutations in Cancer

Teresa Kaserer and Julian Blagg

Combining mutational signatures, clonal fitness and drug affinity to define drug-specific resistance mutations in cancer

Teresa Kaserer,* Julian Blagg*†

Cancer Research UK Cancer Therapeutics Unit, The Institute of Cancer Research, London, SM2 5NG, UK

* Corresponding authors: teresa.kaserer@icr.ac.uk, julian.blagg@icr.ac.uk

† Lead contact: julian.blagg@icr.ac.uk

Table S1. Proportion of DPMs and Insertions among unique cancer mutations. Related to STAR Methods, Introduction and Results.

| Dataset	DPMs (%)	Insertions (%)	Comments	Reference
IARC P53 Database	4.5	12.1	Version R18	http://p53.iarc.fr/ (Bouaoun et al., 2016)
TRACERx	1.2	0.7	The whole cohort	(Jamal-Hanjani et al., 2017)
TRACERx	1.2	1.0	The relapse and autopsy cohort	(Abbosh et al., 2017)
COSMIC Resistance Mutations	5.3	1.5	Version 83 (access date 05/01/2018) Only mutations with defined genetic alteration.	http://cancer.sanger.ac.uk/cosmic (Bamford et al., 2004)
COSMIC Mutation Data	4.1	7.6	Version 83 (access date 05/01/2018)	http://cancer.sanger.ac.uk/cosmic (Bamford et al., 2004)
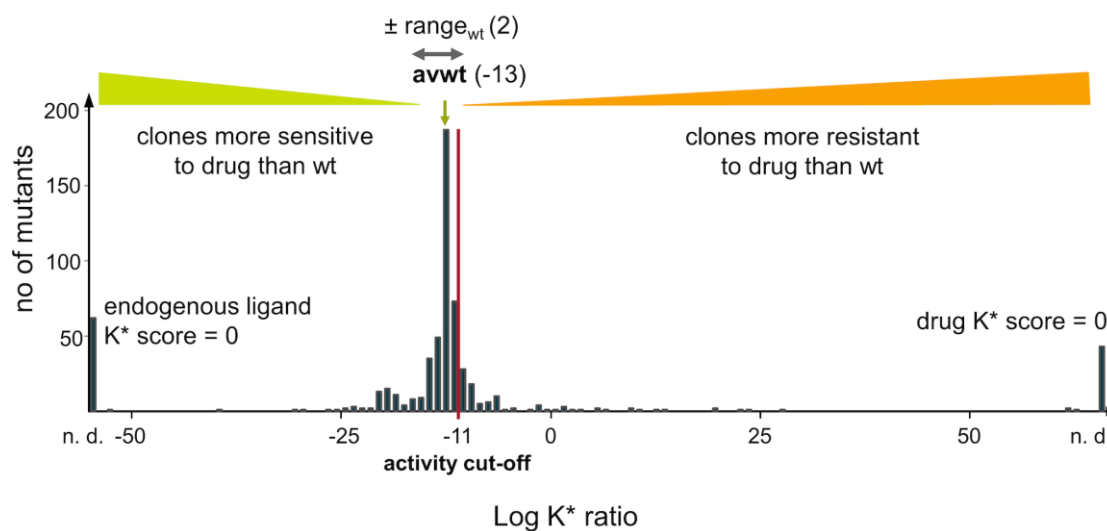
Figure S1. Log K* ratio distribution exemplified for the drug-target pair SCH772489-ERK2. Data related to Discussion and STAR Methods. The number of mutants with a similar K* ratio is plotted against the binned log K* ratio value. Mutants with a score lower than average wt (avwt) – range$_{wt}$ indicate increased drug binding (=sensitivity) until the case where a mutant can still interact with the drug, but not the endogenous ligand (endogenous ligand K* score = 0). Scores higher than the avwt + range$_{wt}$ indicate mutants with increasing resistance to drug, until the case where drug binding is completely abrogated (drug K* score = 0). In the SCH772489-ERK2 example, the logK* score of the avwt is -13 and the wt values span across 2 log-units (= range$_{wt}$). The cut-off was therefore defined as -11. n. d. not defined, when a K* score of 0 was retrieved, the log value could not be defined. The Figure was generated with DataWarrior (Sander et al., 2015).
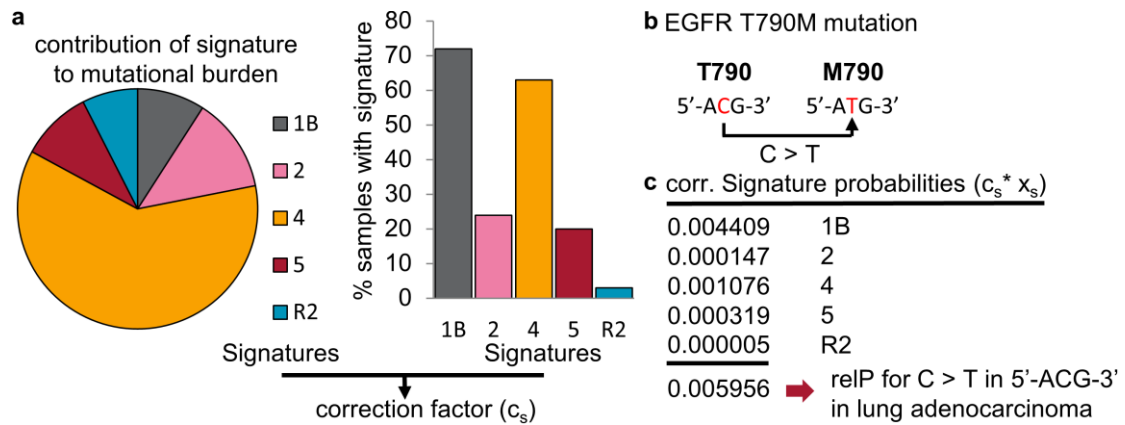
Figure S2. Calculation of relPs exemplified for EGFR T790M in lung adenocarcinoma. Data related to Star Methods. (**a**) Both the contribution of a signature to the mutational burden observed in a particular cancer type, lung adenocarcinoma in this example, and the percentage of patient samples in which the signature was found to be operating, were used to calculate a correction factor (c_s). Signature probabilities reported by Alexandrov et al. (Alexandrov et al., 2013) were multiplied by c_s to normalize signatures for their contribution to lung adenocarcinoma (c_s*x_s). For example, Signature 1B contributes 9.1% of mutations observed in lung adenocarcinoma and was detected in 72% of cancer samples. The c_{1B} for Signature 1B (0.091 x 0.72 = 0.06552) in the context of lung adenocarcinoma was used to normalize the original Signature 1B probabilities (x_B) (Alexandrov et al., 2013). Besides Signature 1B, Signatures 2, 4, 5, and R2 were also reported to contribute to the mutation spectrum in lung adenocarcinoma. Taken together the corrected signature probabilities reflect the likelihood that a particular mutation will occur in that cancer type (pSPM). (**b**) The EGFR T790M mutation is generated by a C > T base change in the context of a 5'-ACG-3' sequence. (**c**) The corrected (corr.) signature probabilities (c_s*x_s) for C > T mutation in the 5'-ACG-3' context for all relevant mutational signatures were added to give the relP for the T790M mutation in lung adenocarcinoma.

Please note, that in this case pSMP = relP, because there is only one base change which can facilitate the T790M mutation. Figure S2a was adapted with permission from Springer Nature: Nature (Alexandrov et al., 2013), COPYRIGHT 2013.