Mapping kinase domain resistance mechanisms for the MET receptor tyrosine kinase via deep mutational scanning

Gabriella O. Estevam[1,2], Edmond M. Linossi[3,4], Jingyou Rao[5], Christian B. Macdonald[1], Ashraya Ravikumar[1], Karson M. Chrispens[1,6], John A. Capra[7], Willow Coyote-Maestas[1,8], Harold Pimentel[5,9,10], Eric A. Collisson[11,12], Natalia Jura[3,4,8], James S. Fraser[1,8,*]

Affiliations

1. Department of Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA, United States
2. Tetrad Graduate Program, UCSF, San Francisco, CA, United States
3. Cardiovascular Research Institute, UCSF, San Francisco, CA, United States
4. Department of Cellular and Molecular Pharmacology, UCSF, San Francisco, CA, United States
5. Department of Computer Science, UCLA, Los Angeles, CA, United States
6. Biophysics Graduate Program, UCSF, San Francisco, CA, United States
7. Bakar Computational Health Sciences Institute and Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA, United States
8. Quantitative Biosciences Institute, UCSF, San Francisco, CA, United States
9. Department of Computational Medicine and Human Genetics, UCLA, Los Angeles, CA, United States
10. Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA, United States
11. Human Biology, Fred Hutchinson Cancer Center, Seattle, Washington, United States
12. Department of Medicine, University of Washington, Seattle, Washington, United States

* - jfraser@fraserlab.com

Abstract

Mutations in the kinase and juxtamembrane domains of the MET Receptor Tyrosine Kinase are responsible for oncogenesis in various cancers and can drive resistance to MET-directed treatments. Determining the most effective inhibitor for each mutational profile is a major challenge for MET-driven cancer treatment in precision medicine. Here, we used a deep mutational scan (DMS) of ~5,764 MET kinase domain variants to profile the growth of each mutation against a panel of 11 inhibitors that are reported to target the MET kinase domain. We identified common resistance sites across type I, type II, and type I ½ inhibitors, unveiled unique resistance and sensitizing mutations for each inhibitor, and validated non-cross-resistant sensitivities for type I and type II inhibitor pairs. We augment a protein language model with biophysical and chemical features to improve the predictive performance for inhibitor-treated datasets. Together, our study demonstrates a pooled experimental pipeline for identifying resistance mutations, provides a reference dictionary for mutations that are sensitized to specific therapies, and offers insights for future drug development.

Introduction

Receptor Tyrosine Kinases (RTKs) are critical signaling molecules that activate and regulate cellular pathways. Disruption of typical RTK regulatory mechanisms through point mutations, gene amplification, protein fusions, or autocrine loops can drive the development, maintenance, and spread of cancers. Small molecule inhibitors are designed to disrupt aberrant signaling cascades by selectively targeting the kinase domain, with tyrosine kinase inhibitors (TKIs) like imatinib showing durable treatment outcomes **(Cohen et al., 2021; Attwood et al., 2021)**. Inhibitors are generally designed against either a wild type kinase or a specific mutational profile, yet acquired mutations can alter sensitivity to different inhibitors and undermine efficacy. The most extreme case of this is resistance, which emerges in the treatment of many cancers by TKI selective pressure **(Cohen et al., 2021; Attwood et al., 2021).** These mutations may act by altering kinase stability, expression, conformation, or activity of the target kinase. Although several recurrent mutations at inhibitor-interacting positions have predictable resistance, specific and rare resistance mutations can be associated with the interactions or conformations unique to certain inhibitors.

An attractive strategy to counter resistance is optimizing the interactions that differ between inhibitors. Small-molecule kinase inhibitors fall into four distinct groups, characterized by their binding modality to the ATP pocket and conformational preferences **(Arter et al., 2022; Attwood et al., 2021; Zuccotto et al., 2009).** Among these groups, three are ATP-competitive: type I, type II, and type I½ **(Figure 1A-C)**. Type I inhibitors occupy the adenosine binding pocket, form hydrogen bonds with "hinge" region residues, and favor an active conformation. Type II inhibitors also occupy the adenosine pocket but extend into an opening in the R-spine that is accessible in an inactive conformation **(Arter et al., 2022) (Figure 1B)**. Type I½ inhibitors combine features from both type I and type II inhibitors, engaging with both the adenosine pocket and the R-spine pocket **(Arter et al., 2022) (Figure 1B)**. Finally, type III inhibitors are allosteric, non-ATP competitive inhibitors **(Arter et al., 2022) (Figure 1B)**. Given the chemical interaction differences and conformational preferences among TKI groups for distinct kinase states, a general approach to combating resistance is sequential treatment of type I and II inhibitors **(Recondo et al., 2018)**. However, without understanding the potential sensitivity of an acquired resistance mutation to the subsequent inhibitor, the efficacy of such strategies is not guaranteed.

The problem of which inhibitor to use and in what order is exemplified by the choice of inhibitors targeting MET kinase (**Bahcall et al., 2016; Recondo et al., 2020; Fernandes et al., 2021**). MET is an RTK and proto-oncogene that has been implicated in the pathogenesis of gastric, renal, colorectal, and lung cancers **Frampton et al., 2015; Duplaquet et al., 2018; Wood et al., 2021; Lu et al., 2017)**. Molecular profiling and next-generation sequencing of tumor samples has provided insight on cancer-associated MET variants **(Frampton et al., 2015; Bahcall et al., 2021)**. Clinical reports following post-treatment outcomes have documented recurrent resistance mutations at positions such as D1228, Y1230, G1163, L1195, H1094 for MET **(Fernandes et al., 2021; Lu et al., 2017; Recondo et al., 2020)**. The challenge of acquired resistance following MET inhibitor therapy has been approached with strategies including sequential treatment of type I and type II TKIs **(Bahcall et al., 2016; Recondo et al., 2020)**, and combination therapy with type I and type II TKIs **(Bahcall et al., 2022; Fernandes et al., 2021; Smyth et al., 2022)**. However, without extensive documentation of the behavior of resistance and sensitizing mutations in MET towards specific TKIs, there remains a barrier towards leveraging inhibitors for specific mutational responses, optimizing inhibitor pairings, and informing rational drug design. Thousands of compounds have been screened against the MET kinase domain, and while several have undergone clinical trials, currently four MET inhibitors have received FDA approval: crizotinib, cabozantinib, tepotinib, and capmatinib **(Santarpia et al., 2021; Wang and Lu, 2023)**. Nevertheless, the emergence of resistance not only limits the efficacy of these drugs but also poses challenges for second-line therapeutic strategies, particularly in the context of rare and novel mutations.

Previously, we used Deep Mutational Scanning (DMS), a pooled cellular selection experiment, to massively screen a library of nearly all possible MET kinase domain mutations. The juxtamembrane domain is encoded by Exon 14 in MET, serves an incompletely understood negative regulatory function in the kinase **(Ma et al., 2003)** and is recurrently excluded in cancer by somatically encoded exon skipping mutations. By testing this library in the context of a wild-type intracellular domain and the recurrent cancer exon 14 skipped variant (METΔEx14) **(Figure 1E)**, we identified conserved regulatory motifs, interactions involving the juxtamembrane and αC-helix, a critical β5 motif, clinically documented cancer mutations, and classified variants of unknown significance **(Estevam et al., 2024)**. Understanding how these variants respond to specific inhibitors can inform therapeutic strategies, with precedent in inhibitor-based DMS studies across kinases such as ERK, CDK4/6, Src, EGFR, and others **(Brenan et al., 2016; Persky et al., 2020; Chakraborty et al., 2023; An et al., 2023)**.

Here, we explore the landscape of TKI resistance of the MET kinase domain against a panel of 11 inhibitors, utilizing our previously established platform **(Estevam et al., 2024)**. By profiling a near-comprehensive library of kinase domain variants in the MET and METΔEx14 intracellular domain, we captured a diverse range of effects based on inhibitor chemistry and 'type' classifications **(Figure 1)**. Within our screen, mutations that confer resistance and offer differential sensitivities across inhibitors were identified, which can be leveraged in sequential or combination therapy. We use Rosace, a Bayesian fitness scoring framework, to reduce false discovery rates in mutational scoring and allow for post-processing normalization of inhibitor treatments **(Rao et al., 2024)**. With our dataset, we have analyzed differential sensitivities to inhibitor pairs and provided a platform for assessing inhibitor efficacy based on mutational sensitivity and likelihood. Lastly, we augment a protein language model **(Rives et al., 2021; Brandes et al., 2023; Chen and Guestrin, 2016)** with biophysical and chemical features to improve predictions for MET inhibitor datasets, and in the future more effectively learn and predict mutational fitness towards novel inhibitors.

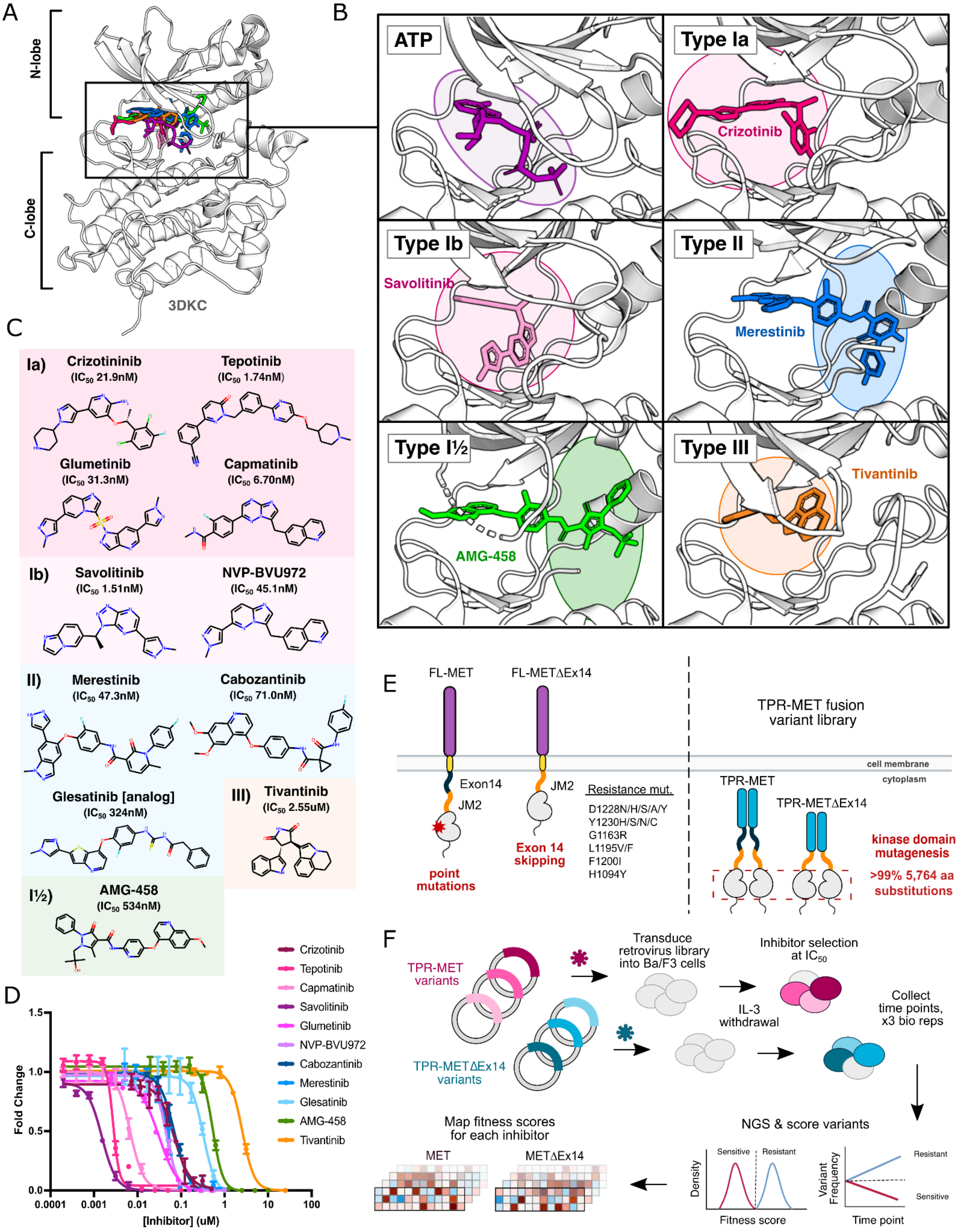

A
N-lobe
C-lobe
3DKC
B
ATP
Type Ia
Crizotinib
Type Ib
Savolitinib
Type II
Merestinib
Type I½
AMG-458
Type III
Tivantinib
C
Ia)
Crizotininib
(IC$_{50}$ 21.9nM)
Tepotinib
(IC$_{50}$ 1.74nM)
Glumetinib
(IC$_{50}$ 31.3nM)
Capmatinib
(IC$_{50}$ 6.70nM)
Ib)
Savolitinib
(IC$_{50}$ 1.51nM)
NVP-BVU972
(IC$_{50}$ 45.1nM)
II)
Merestinib
(IC$_{50}$ 47.3nM)
Cabozantinib
(IC$_{50}$ 71.0nM)
Glesatinib [analog]
(IC$_{50}$ 324nM)
III)
Tivantinib
(IC$_{50}$ 2.55uM)
I½)
AMG-458
(IC$_{50}$ 534nM)
D
Fold Change
1.5
1.0
0.5
0.0
0.0001
0.001
0.01
0.1
1
10
100
[Inhibitor] (uM)
Crizotinib
Tepotinib
Capmatinib
Savolitinib
Glumetinib
NVP-BVU972
Cabozantinib
Merestinib
Glesatinib
AMG-458
Tivantinib
E
FL-MET
FL-METΔEx14
Exon14
JM2
point mutations
JM2
Exon 14 skipping
Resistance mut.
D1228N/H/S/A/Y
Y1230H/S/N/C
G1163R
L1195V/F
F1200I
H1094Y
TPR-MET fusion variant library
cell membrane
cytoplasm
TPR-MET
TPR-METΔEx14
kinase domain mutagenesis
>99% 5,764 aa substitutions
F
TPR-MET variants
TPR-METΔEx14 variants
Transduce retrovirus library into Ba/F3 cells
IL-3 withdrawal
Inhibitor selection at IC$_{50}$
Collect time points, x3 bio reps
NGS & score variants
Density
Sensitive
Resistant
Fitness score
Variant Frequency
Resistant
Sensitive
Time point
Map fitness scores for each inhibitor
MET
METΔEx14

Figure 1. MET kinase inhibitor types and resistance mutations screened against a nearly comprehensive library of kinase domain substitutions. (A) Crystal structure of the ATP-bound MET kinase domain (3DKC) overlaid with type Ia (crizotinib, 2WGJ), type Ib (savolitinib, 6SDE), type II (merestinib, 4EEV), type I½ (AMG-458, 5T3Q), and type III inhibitors (tivantinib, 3RHK). (B) Pocket view of ATP and each inhibitor type bound to the active site of the MET kinase domain with the respective inhibitor and crystal structures from panel A. (C) 2D chemical structures of each inhibitor screened against the site saturation mutagenesis library of the MET kinase domain, with each experimentally determined IC_{50} values displayed for Ba/F3 cells stably expressing the wild-type MET ICD in a TPR-fusion background. (D) Dose-response curves for each inhibitor against the wild-type MET intracellular domain expressed in a TPR-fusion in the Ba/F3 cell line. (E) Schematics of the full-length and exon 14 skipped MET receptor alongside the TPR-fusion constructs with the full-length and exon 14 skipped intracellular domain, displaying four mechanisms of oncogenic activity: point mutations, exon 14 skipping, constitutive activity through domain fusions, and inhibitor resistance mutations. (F) Experimental workflow for defining the mutational landscape of the wild-type TPR-MET and exon 14 skipped TPR-METΔEx14 intracellular domain against 11 ATP-competitive inhibitors in Ba/F3, interleukin-3 (IL-3) withdrawn pooled competition assay.

Results

Measuring the mutational fitness of 5,764 MET kinase domain variants against ATP-competitive inhibitors

To evaluate the response of MET mutations to different inhibitors, we selected six type I inhibitors (crizotinib, capmatinib, tepotinib, glumetinib, savolitinib, and NVP-BVU972), three type II inhibitors (cabozantinib, glesatinib analog, merestinib), and a proposed type III inhibitor, tivantinib **(Figure 1C)**. Type I MET inhibitors leverage pi-stacking interactions with Y1230 and salt-bridge formation between D1228 and K1110, and are further classified as type Ia or Ib based on whether they interact with solvent front residue G1163 **(Figure 1C) (Cui 2014; Fujino et al., 2019; Wang et al., 2023)**. Here we specifically define type Ia inhibitors as having a solvent-front interaction **(Cui 2014)**, which structurally classifies tepotinib and capmatinib as type Ia based on our analysis and inhibitor docking models **(Figure 1A-C)**, despite classification as type Ib in other studies **(Brazel et al., 2022; Fujino et al., 2022; Recondo et al., 2020) (Figure 1 - figure supplement 1).**

As in our previous work, we employed the Ba/F3 cell line as our selection system due to its undetectable expression of endogenous RTKs and addiction to exogenous interleukin-3 (IL-3) **(Estevam et al., 2024)**. These properties allow for positive selection based on ectopically expressed kinase activity and proliferation in the absence of IL-3 **(Daley and Baltimore, 1988; Warmuth et al., 2007; Koga et al., 2022)**. We used a TPR-MET fusion to generate IL-3 independent constitutive activity **(Estevam et al., 2024)**. While this system affords cytoplasmic expression, constitutive oligomerization, and HGF-independent activation, features like membrane-proximal effects are lost **(Cooper et al., 1984; Park et al., 1986; Peschard et al., 2001; Rodrigues and Park, 1993; Vigna et al., 1999; Mak et al., 2007; Pal et al., 2017; Lu et al., 2017; Fujino et al., 2019**). The constitutive activity of TPR-MET and the reliance of Ba/F3 cells on that activity render this selection system quite sensitive for determining reduction in growth from small molecule inhibition.

We generated dose response curves for each inhibitor against wild-type TPR-MET (wild-type intracellular domain, including exon 14 that contains the CBL docking site) and TPR-METΔEx14 (exon 14 skipped intracellular domain) constructs, stably expressed in Ba/F3 cells to determine IC_{50} values for our system **(Figure 1D; Figure 1 - figure supplement 1).** We used our previously published library harboring >99% of all possible 5,764 kinase domain (1059-1345aa) mutations in a TPR-fusion background carrying either a wild-type MET or exon 14 skipped intracellular domain **(Figure 1E) (Estevam et al., 2024)**. Time points were selected every two cell doublings over the course of four time points, and cells were split and maintained in the absence of IL-3 and drugs at IC_{50} for each inhibitor, including a DMSO control. All samples, across all time points and replicates, were prepared for next-generation sequencing (NGS) in parallel, and sequenced on the same Illumina NovaSeq 6000 flow cell to identify variant frequencies **(Figure 1F)**. We then calculated variant fitness scores using Rosace **(Rao et al., 2024) (Figure 1F; Figure 1 - figure supplement 1-4)**. We performed parallel analysis of the TPR-MET and TPR-METΔEx14 screens; however, we focus our analysis below on TPR-MET with the parallel and largely consistent analyses of TPR-METΔEx14 available in the supplement.

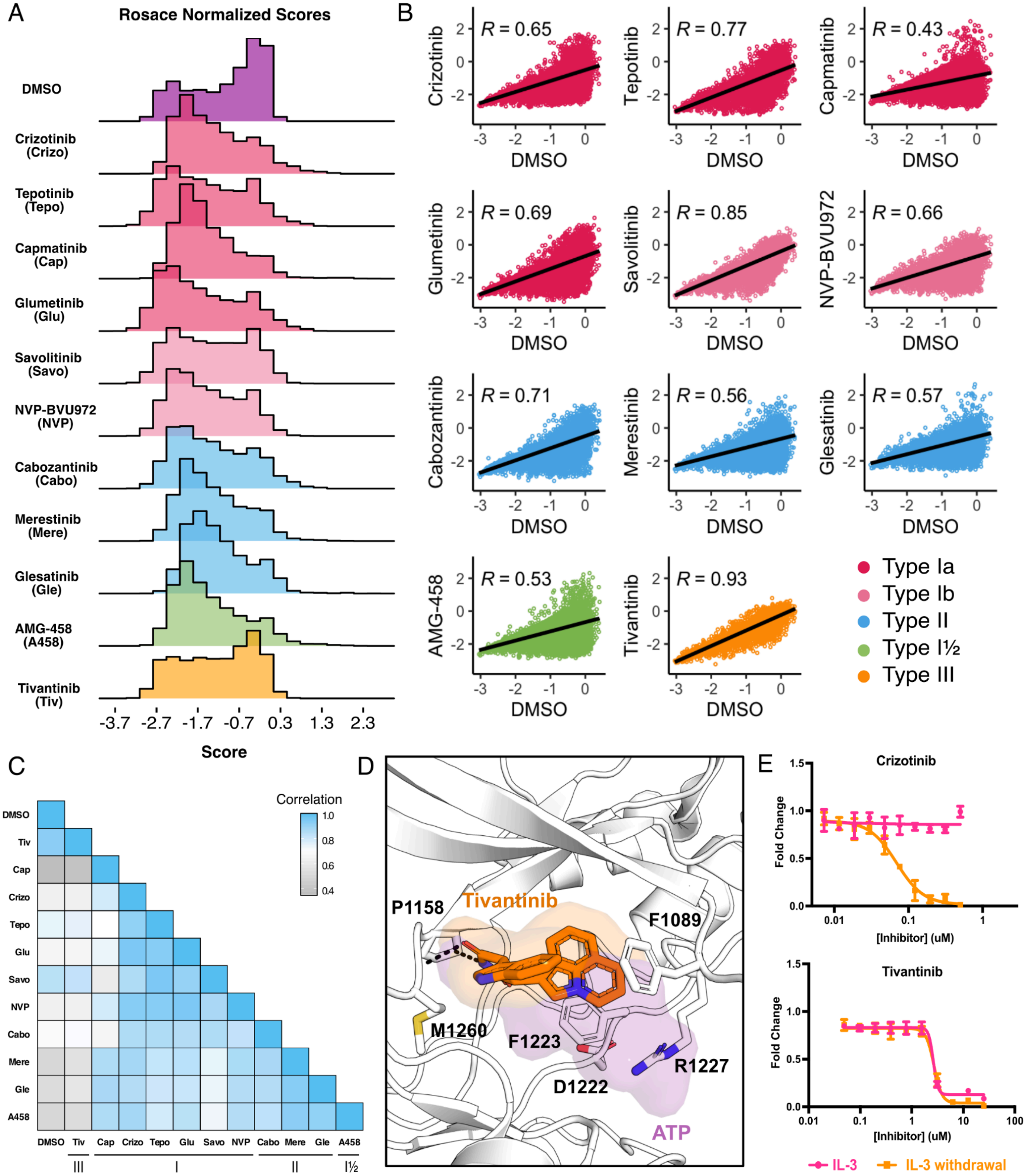

A
Rosace Normalized Scores
DMSO
Crizotinib (Crizo)
Tepotinib (Tepo)
Capmatinib (Cap)
Glumetinib (Glu)
Savolitinib (Savo)
NVP-BVU972 (NVP)
Cabozantinib (Cabo)
Merestinib (Mere)
Glesatinib (Gle)
AMG-458 (A458)
Tivantinib (Tiv)
-3.7 -2.7 -1.7 -0.7 0.3 1.3 2.3
Score
B
Crizotinib R = 0.65
Tepotinib R = 0.77
Capmatinib R = 0.43
Glumetinib R = 0.69
Savolitinib R = 0.85
NVP-BVU972 R = 0.66
Cabozantinib R = 0.71
Merestinib R = 0.56
Glesatinib R = 0.57
AMG-458 R = 0.53
Tivantinib R = 0.93
DMSO
-3 -2 -1 0
2 0 -2
Type Ia
Type Ib
Type II
Type I½
Type III
C
Correlation
1.0 0.8 0.6 0.4
DMSO
Tiv
Cap
Crizo
Tepo
Glu
Savo
NVP
Cabo
Mere
Gle
A458
III
I
II
I½
D
Tivantinib
P1158
F1089
M1260
F1223
D1222
R1227
ATP
E
Crizotinib
Tivantinib
Fold Change
1.5 1.0 0.5 0.0
0.01 0.1 1
0.01 0.1 1 10 100
[Inhibitor] (uM)
IL-3
IL-3 withdrawal

Figure 2. Mutational landscape of the MET kinase domain under 11 ATP-competitive inhibitor selection. (A) Distributions of all variants (wild-type synonymous, early stop, and missense) for each condition in the wild-type TPR-MET kinase domain, scored with Rosace and normalized to the growth rate of the DMSO control population. (B) Correlation plots for all mutational fitness scores for each drug against DMSO, fitted with a linear regression and Pearson's R value displayed. (C) Heatmap showing the Pearson's R correlation for each condition against each other, annotated by condition and inhibitor type. Correlations are colored according to a scale bar from gray to blue (low to high correlation). (D) Crystal structure of the tivantinib-bound MET kinase domain (PDB 3RHK) overlaid with the ATP-bound kinase domain (PDB 3DKC), with tivantinib-stabilizing residues and overlapping density of tivantinib (orange) and ATP (purple) highlighted. (E) Dose responses of crizotinib and tivantinib tested against stable Ba/F3 cells expressing the wild-type intracellular domain of MET fused to TPR, tested in the presence and absence of interleukin-3 (IL-3).

Defining the mutational landscape of resistant and sensitizing mutations for the MET kinase domain.

While growth rates were experimentally controlled through equipotent dosing during selection, there was no direct way to validate this post-processing. To generate meaningful comparisons between inhibitor scores and conditions, in addition to performing downstream score subtractions, we normalized cell growth rates for each inhibitor to the growth rate observed for the DMSO population. As expected, the DMSO control population displayed a bimodal distribution with mutations exhibiting wild-type fitness centered around 0, with a wider distribution of mutations that exhibited loss- or gain-of-function effects **(Figure 2A).** Also as expected, inhibitor-treated populations displayed distributions with a loss-of-function peak, representative of mutations that are sensitive to the inhibitor. Unlike DMSO, inhibitor populations were right-skewed, showing greater enrichment of gain-of-function scores at the positive tail of distributions **(Figure 2A).** These population differences were exemplified by the low correlation between each inhibitor and DMSO, with capmatinib showing the greatest difference from DMSO with a Pearson's correlation of 0.45, and tivantinib standing as an outlier with a Pearson's correlation of 0.93 **(Figure 2B).**

In comparing all conditions to each other, we were able to further capture differences between and within inhibitor types. Type II inhibitors displayed the greatest similarities to one another, with merestinib and the glesatinib analog having the highest correlation (r = 0.93) and cabozantinib and glesatinib analog showing the lowest (r = 0.87) **(Figure 2C)**. While type I inhibitors were also highly correlated, capmatinib stood out as an outlier, displaying the lowest correlations potentially due to difficulty in experimental overdosing due to its greater potency **(Bahcall et al., 2022; Fujino et al., 2019**) **(Figure 2C)**. While there was only one type I½ inhibitor, AMG-458, it displayed higher similarity to type II inhibitors than to type I, likely due to similar type II back pocket interactions with the kinase R-spine **(Figure 1B; Figure 2C)**. Nevertheless, AMG-458 was most distinct from cabozantinib (r = 0.83) and type I inhibitors tepotinib (r = 0.79) and savolitinib (r = 0.73) **(Figure 2C)**. Between type I and type II groups, with the exception of capmatinib, tepotinib and savolitinib showed the lowest correlation with merestinib and glesatinib analog **(Figure 2C)**.

The strong correlation of tivantinib with the DMSO control (r = 0.93) and low correlation with all other inhibitors suggested a MET-independent mode of action. Until recently **(Michaelides et al, 2023)**, tivantinib was considered the only type III MET-inhibitor and showed promising early clinical trial results **(Eathiraj et**

al., 2011; Wang et al., 2022). *In vitro* assays on the purified MET kinase domain have shown that tivantinib has the potential to hinder catalytic activity **(Munshi et al., 2010)** and structural studies revealed that it selectively targets an inactive DFG-motif conformation, with tivantinib stabilizing residues (F1089, R1227) blocking ATP binding **(Figure 2D) (Eathiraj et al., 2011)**. Yet in contradiction, comparative MET-dependent and MET-independent cell-based studies on tivantinib have also shown MET agnostic anti-tumor activity, posing that tivantinib may have an alternative inhibitory mechanism than a MET-selective one **(Michieli and Di Nicolantonio, 2013; Basilico et al., 2013; Katayama et al., 2013; Fujino et al., 2019)**.

Therefore, to test the hypothesis that tivantinib is not MET-selective in our system, we compared the dose response of tivantinib and crizotinib in the presence and absence of IL-3 for wild-type TPR-MET, stably expressed in Ba/F3 cells **(Figure 2E)**. As expected, crizotinib only displayed an inhibitory effect under IL-3 withdrawal, highlighting a MET-dependent mode of action. In contrast, tivantinib displayed equivalent inhibition regardless of IL-3, reinforcing that tivantinib has cytotoxicity effects unrelated to MET inhibition **(Figure 2E)** and underscores the sensitivity of the DMS in identifying direct protein-drug effects.

Crizotinib-MET kinase domain resistance profiles exemplify the information accessible from individual inhibitor DMS

As an example of the insights that can be learned from the inhibitor DMS screens, we examined the profile for crizotinib, one of four FDA approved inhibitors for MET and a multitarget TKI **(Cui et al., 2011; Wang and Lu, 2023; Santarpia et al., 2021).** To identify mutations that show gain-of-function and loss-of-function behaviors specific to inhibitors compared to DMSO, we subtracted DMSO from all fitness scores. **(Figure 3A),** with the expectation that effects related to expression or stability would be similar in both conditions, enhancing the ability to identify drug sensitivity or resistance. Indeed, the highest frequency of gain-of-function mutations occurred at residues mediating direct drug-protein interactions, such as D1228, Y1230, and G1163. These sites, and many of the individual mutations, have been noted previously in clinical reports **(Fernandes et al., 2021; Yao et al., 2023; Tovar and Graveel, 2017; Tiedt et al., 2011)**. Beyond these well characterized sites, regions with sensitivity occurred throughout the kinase, primarily in loop-regions which have the greatest mutational tolerance in DMSO, but do not provide a growth advantage in the presence of an inhibitor **(Figure 1 - figure supplement 1; Figure 1 - figure supplement 2)**.

In mapping positions with resistance to the crizotinib-bound kinase domain crystal structure (PDB 2WGJ), our DMS results further emphasize the emergence of resistance mutations at the ATP-binding site and direct-protein drug interacting residues **(Figure 3B-D)**. Mutations to the hinge position, Y1159, and C-spine residues, including M1211 and V1092, introduce charge or are predicted to change the conformation of the pocket to clash with crizotinib but not ATP **(Figure 3E-G)**. Outside of direct drug-protein interactions, positions I1084, T1261, Y1093, and G1242 displayed the largest resistance signals **(Figure 3A-D)**. Structurally, I1084 is located in β1 at the roof of the ATP-binding pocket. The R-group of a His mutation at this position clashes with crizotinib's hinge-binding and solvent-front moieties but does not interfere with bound-ATP **(Figure 3E-F).** Y1093 is also at the roof of the ATP-binding site, residing in β2 **(Figure 3B)**. However, its structural influence on resistance is unobvious. In all rotameric states, the R-group of Y1093 points away from the catalytic site and does not clash with crizotinib. We speculate that Y1093 mutations may negatively impact crizotinib's stability in the catalytic site compared to ATP, as ATP's triphosphate groups are stabilized by the P-loop. Therefore, comparing crizotinib to DMSO highlights both known hotspots and rarer sites like I1084 and Y1093, which may contribute to resistance through conformational changes rather than disrupting direct inhibitor-protein contacts. Specific inhibitor fitness landscapes also aid in identifying mutations with potential drug sensitivity, such as R1086 and C1091 in the MET P-loop, which might represent attractive interacting sites for further drug design with a lower probability of acquiring resistance.

We identified unique resistance mutations enriched at the ATP-binding site across all inhibitors, yet also noticed discernible differences between type I and II inhibitors, the R-spine, and αC-helix **(Figure 4A-G)**. Mapping inhibitor-specific positions and mutational scores, not only provides a mutation-level breakdown of inhibitor contributions to common resistance mutations, but also demonstrates differences in structural resistance enrichment across specific inhibitors **(Figure 4A-G)**. To summarize this information, we next examined trends by grouping inhibitors by type.

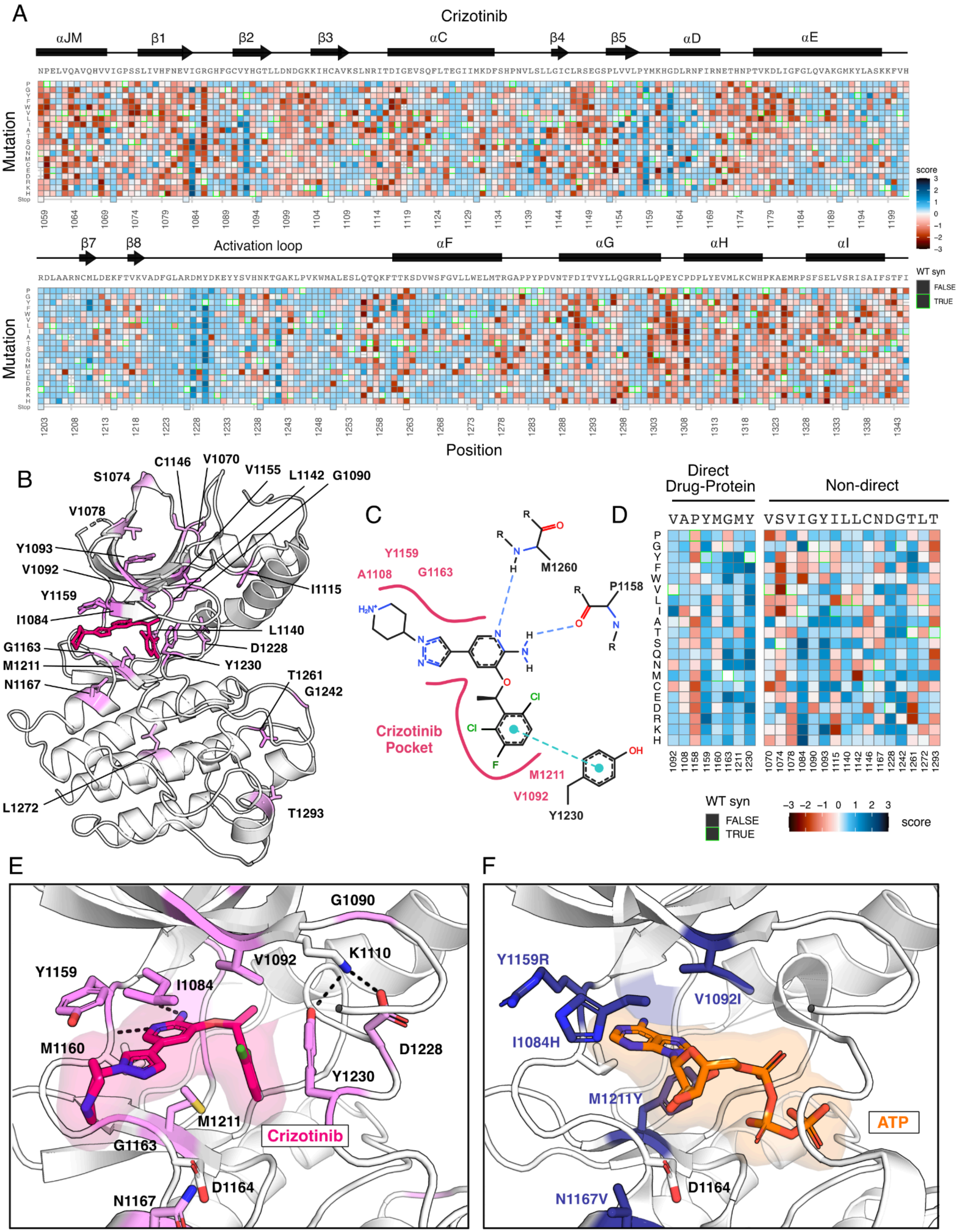

A
Crizotinib
αJM
β1
β2
β3
αC
β4
β5
αD
αE
NPELVQAVQHVVIGPSSLIVHFNEVIGRGHFGCVYHGTLLDNDGKKIHCAVKSLNRITDIGEVSQFLTEGIIMKDFSHPNVLSLLGICLRSEGSPLVVLPYMKHGDLRNFIRNETHNPTVKDLIGFGLQVAKGMKYLASKKFVH
Mutation
P G Y F W V L I A T S Q N M C E D R K H Stop
1059 1064 1069 1074 1079 1084 1089 1094 1099 1104 1109 1114 1119 1124 1129 1134 1139 1144 1149 1154 1159 1164 1169 1174 1179 1184 1189 1194 1199
score
3 2 1 0 −1 −2 −3
β7
β8
Activation loop
αF
αG
αH
αI
RDLAARNCMLDEKFTVKVADFGLARDMYDKEYYSVHNKTGAKLPVKWMALESLQTQKFTTKSDVWSFGVLLWELMTRGAPPYPDVNTFDITVYLLQGRRLLQPEYCPDPLYEVMLKCWHPKAEMRPSFSELVSRISAIFSTFI
WT syn
FALSE
TRUE
1203 1208 1213 1218 1223 1228 1233 1238 1243 1248 1253 1258 1263 1268 1273 1278 1283 1288 1293 1298 1303 1308 1313 1318 1323 1328 1333 1338 1343
Position
B
S1074
C1146
V1070
V1155
L1142
G1090
V1078
Y1093
V1092
Y1159
I1084
G1163
M1211
N1167
I1115
L1140
D1228
Y1230
T1261
G1242
L1272
T1293
C
Y1159
A1108
G1163
M1260
P1158
Crizotinib Pocket
M1211
V1092
Y1230
D
Direct Drug-Protein
Non-direct
VAPYMGMY
VSVIGYILLCNDGTLT
1092 1108 1158 1159 1160 1163 1211 1230
1070 1074 1078 1084 1090 1093 1115 1140 1142 1146 1167 1228 1242 1261 1272 1293
WT syn
FALSE
TRUE
score
E
G1090
Y1159
V1092
K1110
I1084
M1160
D1228
Y1230
M1211
Crizotinib
G1163
D1164
N1167
F
Y1159R
V1092I
I1084H
M1211Y
ATP
N1167V
D1164

Figure 3. Novel resistance mutations identified and mapped for crizotinib. (A) Heatmap of crizotinib fitness scores subtracted from DMSO, scaled from loss-of-function (red) to gain-of-function (blue), with the wild-type protein sequence, secondary structure, kinase domain residue position, and mutational substitution annotated. Wild-type synonymous substitutions are outlined in green. (B) Resistance positions mapped onto the crizotinib-bound, MET crystal structure (PDB 2WGJ). Positions are labeled and colors are scaled according to the average score for the resistance mutations at each site. (C) 2D protein-drug interactions between crizotinib and the MET kinase domain (PDB 2WGJ) with pocket residues and polar and pi interactions annotated. Schematic generated through PoseEdit **(Diedrich et al., 2023)** (https://proteins.plus/). (D) Condensed crizotinib heatmap displaying direct drug-protein interacting and non-direct resistance position. (E) Crizotinib binding site and pocket residues displayed with resistance positions highlighted (pink) and the wild-type residue and inhibitor interactions shown (PDB 2WGJ). (F) Resistance mutations modeled for I1084H, V1092I, Y1159R, M1211Y, and N1167 relative to ATP (PDB 3DKC).

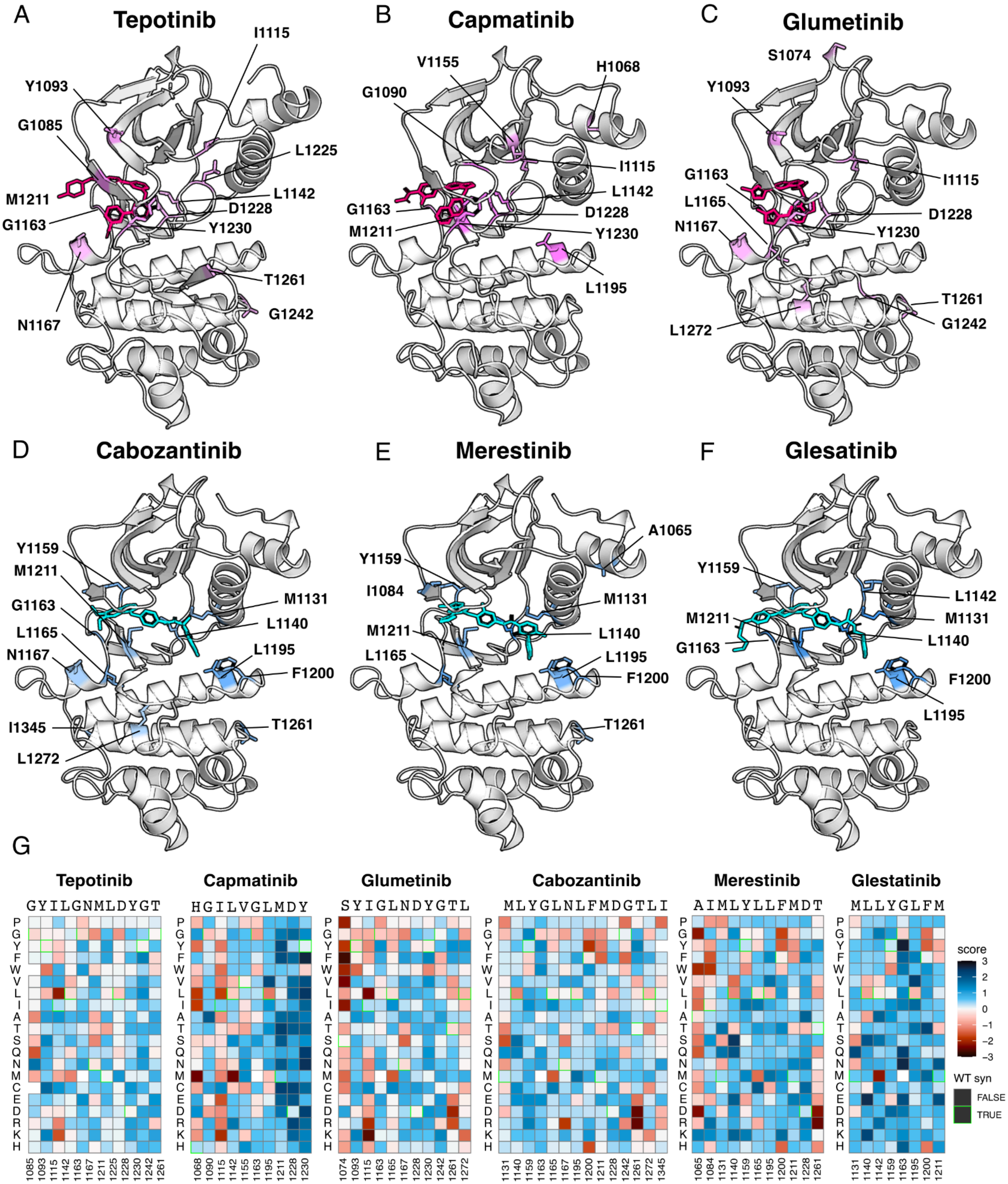

A
Tepotinib
I1115
Y1093
G1085
L1225
M1211
L1142
D1228
G1163
Y1230
T1261
G1242
N1167
B
Capmatinib
V1155
H1068
G1090
I1115
L1142
G1163
D1228
M1211
Y1230
L1195
C
Glumetinib
S1074
Y1093
I1115
G1163
L1165
D1228
N1167
Y1230
T1261
G1242
L1272
D
Cabozantinib
Y1159
M1211
G1163
M1131
L1140
L1165
N1167
L1195
F1200
I1345
T1261
L1272
E
Merestinib
A1065
Y1159
I1084
M1131
M1211
L1140
L1165
L1195
F1200
T1261
F
Glesatinib
Y1159
L1142
M1211
M1131
G1163
L1140
F1200
L1195
G
Tepotinib
GYILGNMLDYGT
1085 1093 1115 1142 1163 1167 1211 1225 1228 1230 1242 1261
Capmatinib
HGILVGLMDY
1068 1090 1115 1142 1155 1163 1195 1211 1228 1230
Glumetinib
SYIGLNDYGTL
1074 1093 1115 1163 1165 1167 1228 1230 1242 1261 1272
Cabozantinib
MLYGLNLFMDGTLI
1131 1140 1159 1163 1165 1167 1195 1200 1211 1228 1242 1261 1272 1345
Merestinib
AIMLYLLFMDT
1065 1084 1131 1140 1159 1165 1195 1200 1211 1228 1261
Glestatinib
MLLYGLFM
1131 1140 1142 1159 1163 1195 1200 1211
P G Y F W V L I A T S Q N M C E D R K H
score
3
2
1
0
−1
−2
−3
WT syn
FALSE
TRUE

Figure 4. Resistance mutations mapped onto experimental and docked kinase domain structures for type I and type II inhibitors. (A-F) Resistance positions and average resistance mutational score mapped onto representative crystal structures (tepotinib, 4R1V; merestinib, 4EEV) and labeled (type I, pink; type II, blue). Inhibitors lacking experimental structures (capmatinib, cabozantinib, glumetinib, and glesatinib) were docked onto a representative type I (PDB 2WGJ) and type II (4EEV) structure through AutoDock Vina **(Eberheart et al., 2021; Trott et al., 2010).** (G) Heatmaps of each resistance position structurally modeled for each drug with fitness scores scaled from loss-of-function (red) to gain-of-function (blue). Wild-type synonymous substitutions are outlined in green.

Novel resistance mutations identified for type I, type II, and type I ½ inhibitors

To reveal patterns across different inhibitor types, we filtered resistance mutations by their score and test statistics **(Figure 5 - figure supplement 1)** and collapsed the information by inhibitor type, plotting the total frequency of resistance mutations at each position **(Figure 5A).** In this condensed heatmap, several common resistance positions emerged within and across inhibitor types to provide a broad view of “hotspots.” Two positions stood out with the highest frequency of resistance: G1163 and D1228 **(Figure 5A-D)**. Both sites are unsurprising due to their inhibitor interactions - G1163 is at the solvent front entrance of the active site and D1228 stabilizes an inactive conformation of the A-loop with an inhibitor bound **(Cui 2014; Recondo et al., 2020; Fernandes et al., 2021).** Located at the base of the active site, M1211 is a previously documented resistance site **(Tidet et al., 2011)** and a C-spine residue **(Estevam et al., 2024)**, which harbors a smaller number of resistance mutations for all inhibitor types within our analysis **(Figure 5A)**. In contrast to these universal sites, Y1230 was a hotspot for type I and I ½ inhibitors, but not a major resistance site for type II inhibitors **(Figure 5A-D).** This specificity can be rationalized based on the role of Y1230 in stabilizing inhibitors through pi-stacking interactions **(Cui 2014)**. In contrast, F1200 and L1195 **(Bahcall et al., 2022; Recondo et al., 2020),** are both hotspots for type II but not type I inhibitors **(Figure 5A-C).** Again, this effect can be rationalized structurally: both residues make direct contact with type II inhibitors, but not type I inhibitors.

Across all inhibitor types, there were a total of 17 shared variants with G1163, D1228, and M1211 being the most common **(Figure 5G)**. The overall spatial pattern of mutations for each inhibitor type follows general principles that are expected based on their interactions. For example, I1084 is enriched as a resistance site for type I inhibitors, consistent with previous studies in hereditary papillary renal cell carcinomas **(Guérin et al., 2023).** I1084 is located at the solvent front of the phosphate-loop (P-loop) of the kinase N-lobe **(Figure 1A),** which is responsible for stabilization of the ATP phosphate groups. This region of the kinase is leveraged for interactions with type I, but not type II inhibitors. In contrast, L1142, an R-spine residue, and L1140, which sits at the back of the ATP-binding pocket, are enriched for type II inhibitor resistance, consistent with their spatial locations **(Figure 5E-F)**. Resistance mutations tend to cluster around the catalytic site across all types, but the shared mutations across different inhibitor types did not display a unifying pattern that evokes a simple rule for combining or sequencing inhibitors to counter resistance **(Figure 5H)**.

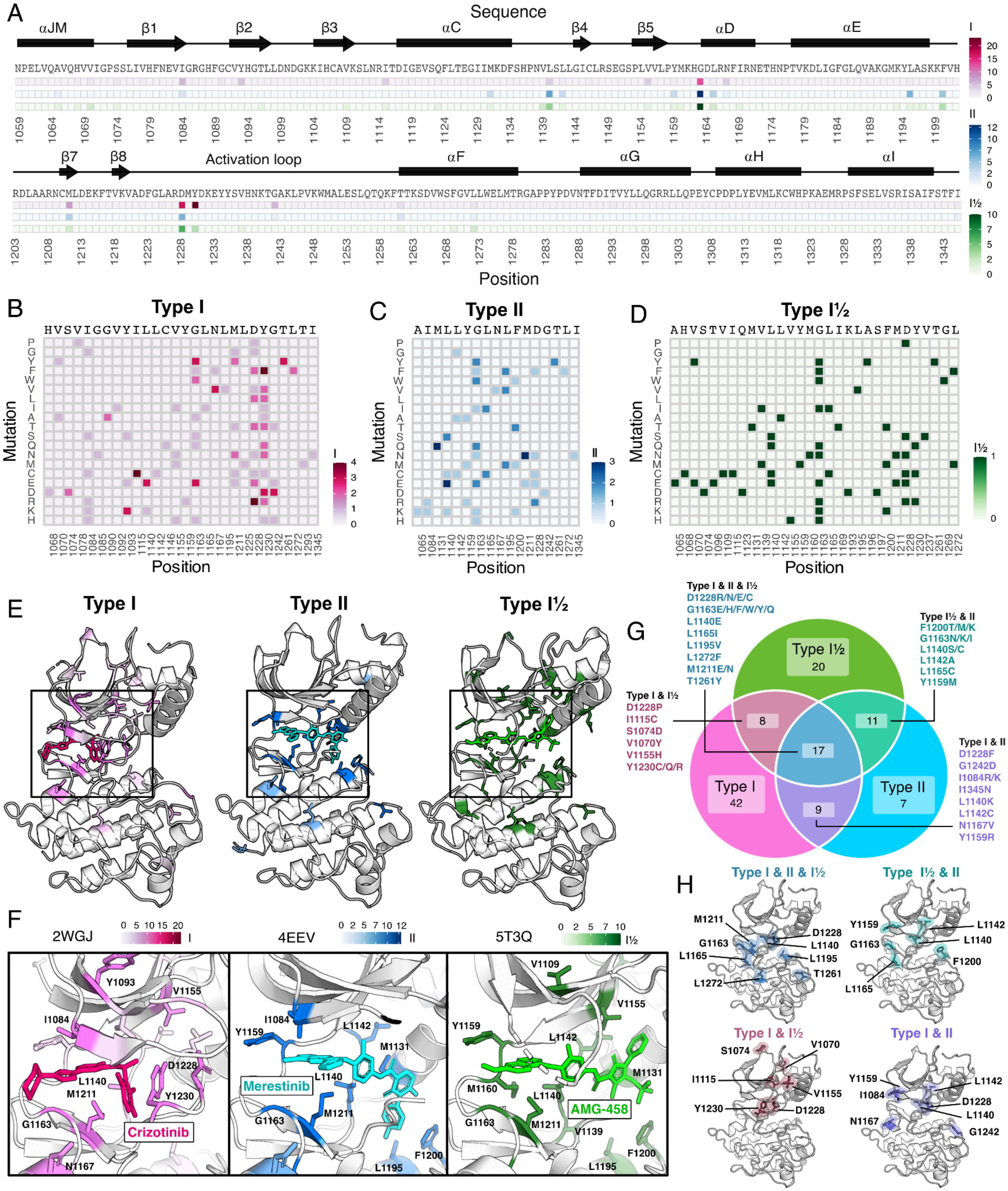

A
Sequence
αJM
β1
β2
β3
αC
β4
β5
αD
αE
NPELVQAVQHVVIGPSSLIVHFNEVIGRGHFGCVYHGTLLDNDGKKIHCAVKSLNRITDIGEVSQFLTEGIIMKDFSHPNVLSLLGICLRSEGSPLVVLPYMKHGDLRNFIRNETHNPTVKDLIGFGLQVAKGMKYLASKKFVH
1059 1064 1069 1074 1079 1084 1089 1094 1099 1104 1109 1114 1119 1124 1129 1134 1139 1144 1149 1154 1159 1164 1169 1174 1179 1184 1189 1194 1199
β7
β8
Activation loop
αF
αG
αH
αI
RDLAARNCMLDEKFTVKVADFGLARDMYDKEYYSVHNKTGAKLPVKWMALESLQTQKFTTKSDVWSFGVLLWELMTRGAPPYPDVNTFDITVYLLQGRRLLQPEYCPDPLYEVMLKCWHPKAEMRPSFSELVSRISAIFSTFI
1203 1208 1213 1218 1223 1228 1233 1238 1243 1248 1253 1258 1263 1268 1273 1278 1283 1288 1293 1298 1303 1308 1313 1318 1323 1328 1333 1338 1343
Position
I
20 15 10 5 0
II
12 10 7 5 2 0
I½
10 7 5 2 0
B
Type I
HVSVIGGVYILLCVYGLNLMLDYGTLTI
Mutation
P G Y F W V L I A T S Q N M C E D R K H
1068 1070 1074 1078 1084 1090 1092 1093 1115 1140 1142 1146 1155 1159 1163 1165 1167 1195 1211 1225 1228 1230 1242 1261 1293 1345
Position
I
4 3 2 1 0
C
Type II
AIMLLYGLNLFMDGTLI
Mutation
1065 1084 1131 1140 1142 1159 1163 1165 1167 1195 1200 1211 1228 1242 1261 1272 1345
Position
II
3 2 1 0
D
Type I½
AHVSTVIQMVLLVYMGLIKLASFMDYVTGL
Mutation
1065 1068 1070 1074 1096 1109 1115 1123 1131 1139 1140 1142 1155 1159 1160 1163 1165 1169 1193 1195 1196 1197 1200 1211 1228 1230 1237 1261 1269 1272
Position
I½
1 0
E
Type I
Type II
Type I½
F
2WGJ
0 5 10 15 20 I
Y1093
V1155
I1084
D1228
L1140
M1211
Y1230
G1163
Crizotinib
N1167
4EEV
0 2 5 7 10 12 II
I1084
Y1159
L1142
M1131
Merestinib
L1140
M1211
G1163
F1200
L1195
5T3Q
0 2 5 7 10 I½
V1109
V1155
Y1159
L1142
M1131
M1160
L1140
AMG-458
G1163
M1211
V1139
F1200
L1195
G
Type I & II & I½
D1228R/N/E/C
G1163E/H/F/W/Y/Q
L1140E
L1165I
L1195V
L1272F
M1211E/N
T1261Y
Type I½ & II
F1200T/M/K
G1163N/K/I
L1140S/C
L1142A
L1165C
Y1159M
Type I & I½
D1228P
I1115C
S1074D
V1070Y
V1155H
Y1230C/Q/R
Type I & II
D1228F
G1242D
I1084R/K
I1345N
L1140K
L1142C
N1167V
Y1159R
Type I½
20
Type I
42
Type II
7
8
11
17
9
H
Type I & II & I½
M1211
D1228
G1163
L1140
L1165
L1195
T1261
L1272
Type I½ & II
Y1159
L1142
L1140
G1163
F1200
L1165
Type I & I½
S1074
V1070
I1115
V1155
Y1230
D1228
Type I & II
Y1159
L1142
I1084
D1228
L1140
N1167
G1242

Figure 5. Novel resistance mutations and "hotspots" identified for MET inhibitor types. (A) Collapsed heatmap of common resistance positions along the kinase domain, with the wild-type protein sequence and secondary structure annotated. Each tile represents a sum of counts for statistically filtered resistance mutations across all inhibitors for type I (pink), type II (blue), and the type I½ inhibitor AMG-458 (green), with the scale bar reflecting counts of resistance mutations across respective inhibitor types. (B-D) Expanded heatmap showing each resistance position and the counts for each specific resistance mutation across all inhibitor types type I (pink), type II (blue), and the type I½ inhibitor AMG-458 (green). Wild-type sequence and variant change are annotated. (E-F) Average frequency of resistance mutations for each mapped on to a representative type I (crizotinib, 2WGJ) and type II (merestinib, 4EEV) crystal structure, alongside the type I½, AMG-458 structure (5T3Q), with associated scale bars. Individual positions with high resistance mutation frequencies are annotated on each structure, with a zoom-in of the bound inhibitor and surrounding resistance sites. (G) Venn diagram showing mutations shared among type I (pink), type II (light blue), and type I½ (green). (H) Structurally mapped (PDB 2WGJ) resistance positions shared among type I, II, I½ (blue-gray), type I and II (purple), type I and I½ (dusty rose), type II and I½ (teal) inhibitors.

Differential sensitivities of the MET kinase domain to type I and type II inhibitors

Strategies aimed at preventing resistance, such as sequential or combination dosing of type I and type II inhibitors, have been explored and offer promise in preventing resistance **(Bahcall et al., 2016; Recondo et al., 2020; Bahcall et al., 2022; Fernandes et al., 2021)**. However, the efficacy of these strategies is limited to the emergence of secondary resistance mutations, and specific inhibitor pairings are further limited to case examples of disparate effects. Using our DMS datasets, we sought to identify inhibitor pairings with the largest divergence in cross-sensitivity. By comparing the fitness landscape of each type I inhibitor to each type II inhibitor, we could again assess inhibitor response likeness based on correlations (**Figure 2A; Figure 6 - figure supplement 1)**. Type I and type II pairs with the highest correlations included capmatinib and glesatinib analog (r = 0.92), suggesting a large overlapping fitness profile, in contrast to pairs with the lower correlations, like savolitinib and merestinib (r = 0.7) **(Figure 6A; Figure 6 - figure supplement 1)**. Overall, cabozantinib maintained the lowest average correlation with all type I inhibitors, making it the most divergent type II inhibitor within our screen, and potentially offering the least overlap in resistance **(Figure 6A; Figure 6 - figure supplement 1)**.

To narrow our characterization of cross-sensitivity, we focused on the inhibitor pair crizotinib and cabozantinib **(Figure 6B; Figure 6 - figure supplement 1).** By statistically filtering mutations that are categorized as gain-of-function in one inhibitor, but loss-of-function in the other, a set of 44 mutations were identified as having crizotinib resistance and cabozantinib sensitivity, and 3 mutations with the opposite profile **(Figure 6B,C)**. Structural mapping of divergent mutations further revealed enrichment at the N-lobe and typical protein-drug interaction sites like Y1230, G1163, and M1211 **(Figure 6B, C)**. While these positions have precedence for resistance, as previously noted, they are also resistance hotspots across all inhibitor types **(Figure 5A)**, where even mutations with differential sensitivities may be insufficient targets to counteract the reemergence of resistance, thus limiting the interchangeability of drugs.

Understanding which mutations have resistance profiles for only type I or type II inhibitors provides better leverage for sequential and combination dosing. To identify such mutations across our dataset, we further filtered variants that met our resistance metrics and were only observed for inhibitors of the same type. In again comparing crizotinib to cabozantinib, Y1093K was a mutation with one of the largest differences between crizotinib and cabozantinib, having a gain-of-function profile for crizotinib and loss-of-function for cabozantinib **(Figure 6B).** Interestingly, Y1093 is located in β2 of the N-lobe, at the roof of the ATP-binding site, and does not directly engage with crizotinib. We speculate this mutation potentially contributes to resistance by perturbing the packing of β1-β2 and altering the conformation of the ATP binding site in a manner that destabilizes crizotinib binding. When comparing the dose-response of Y1093K to the wild-type TPR-MET kinase domain, Y1093K shows a nearly 10-fold shift in crizotinib sensitivity with no difference in cabozantinib sensitivity **(Figure 6D)**. In identifying mutations with the opposite profiles, resistance to cabozantinib and sensitivity to crizotinib, L1195M displayed the greatest differential scores **(Figure 6B).** L1195 is an αE-helix position with previously recorded resistance (L1195V/F), which our analysis further supports as a type II-only resistance hotspot **(Figure 5A)**. Structurally, mutations like Met or Phe at 1195 clash with the fluorophenyl moiety of cabozantinib used to access and stabilize a deep, back pocket of the kinase in an inactive conformation, unlike crizotinib which occupies the solvent front and adenosine binding region of the ATP binding site. In comparing the dose-response of L1195M to the wild-type TPR-MET kinase domain, we find that L1195M is refractory to all concentrations of cabozantinib tested, but still sensitive to crizotinib **(Figure 6D)**. Beyond a type I and type II pairing, such cross-resistance identification can be further applied to identify differential sensitivities within an inhibitor group **(Figure 6 - figure supplement 2)**, which can further expand opportunities for inhibitor-specific sensitivity in therapy and drug design.

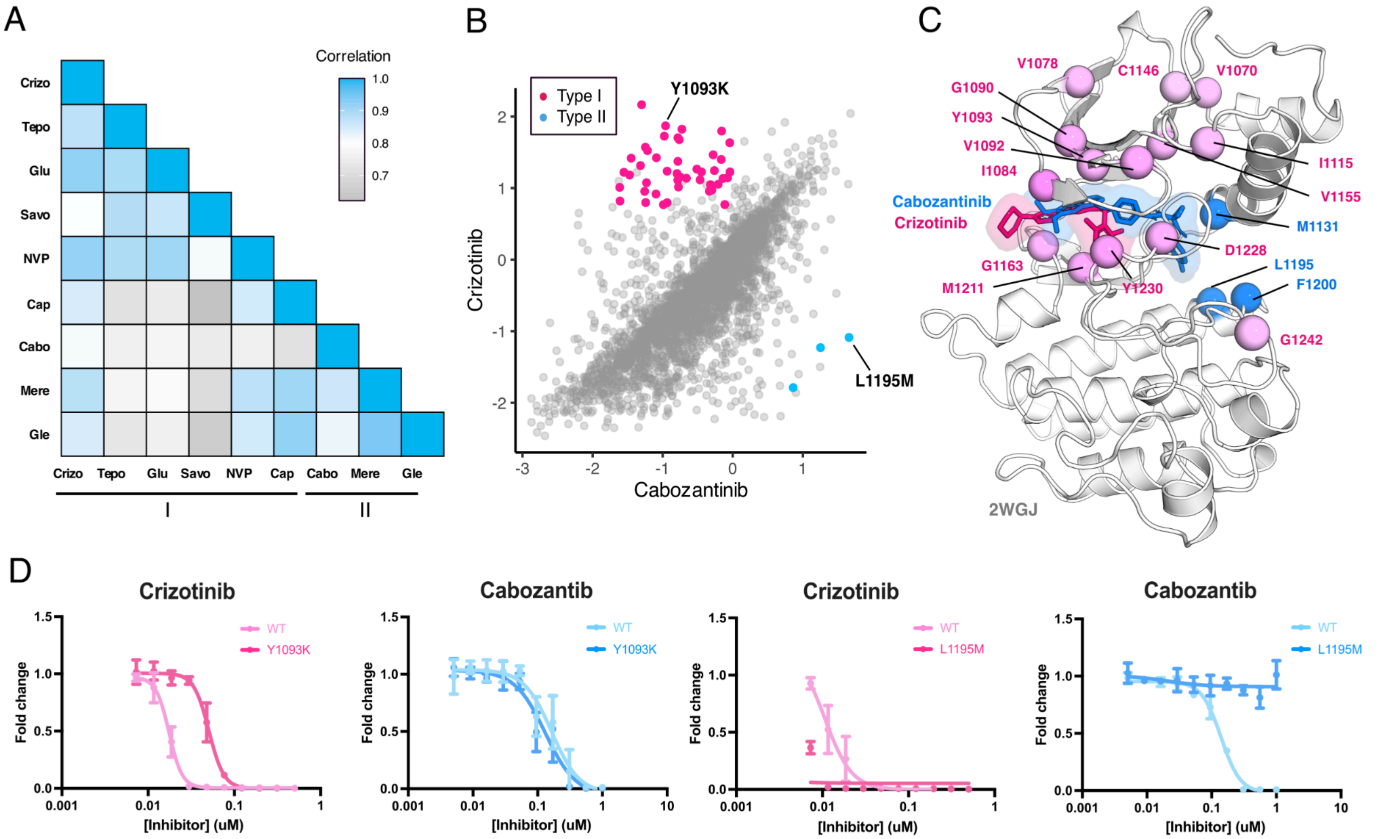


Figure 6. MET kinase domain differential sensitivities revealed for type I and type II inhibitors. (A) Heatmap showing Pearson correlation values for all combinations of screened type I and type II inhibitors. Correlations were determined from DMSO subtracted fitness scores (B) Correlation plot correlation plot of DMSO subtracted fitness scores for crizotinib and cabozantinib. Mutations with differential scores are highlighted for type I (pink) and type II (blue). (C) Average scores of mutations with differential sensitivities within inhibitor pairs mapped and annotated in respective crystal structures (crizotinib, 2WGJ; cabozantinib, docked into 4EEV). Positions that are gain-of-function for type I but loss-of-function in type II are highlighted in pink, whereas positions that are gain-of-function for type II but loss-of-function in type I are highlighted in blue. (D) Dose-response curves for crizotinib and cabozantinib in Ba/F3 cells expressing TPR-MET (full MET intracellular domain) harboring mutations at Y1093K and L1195M.

Identification of biophysical contributors to inhibitor-specific fitness landscapes using machine learning

Machine learning models originally developed for predicting protein structure **(Jumper et al., 2021; Rives et al., 2021; Lin et al., 2023)** have been adapted for predicting protein-ligand complexes **(Bryant et al., 2023)**, and predicting fitness values from DMS studies **(Meier et al., 2021; Brandes et al., 2023; Jones et al., 2020)**. In particular, protein language models have shown the ability to estimate the functional effects of sequence variants in correlation with DMS data **(Rives et al., 2021; Meier et al., 2021; Brandes et al., 2023)**. We observed that ESM-1b, a protein language model **(Rives et al., 2021)**, predicts the fitness of variants in the untreated/DMSO condition (correlation 0.50) much better than it does for the inhibitor treated datasets (correlation 0.28). This difference in predictive ability is likely because the language model is trained on sequences in the evolutionary record and fitness in the presence of inhibitors does not reflect a pressure that has operated on evolutionary timescales.

To overcome this limitation and improve the predictive properties of the ESM approach, we sought to augment the model with additional features that reflect the interactions between protein and inhibitors that are not present in the evolutionary record **(Chen and Guestrin, 2016)**. There are several challenges associated with this task, including the narrow sequence space explored, high correlations between datasets, and the limited chemical space explored by the 11 inhibitors. We used an XGBoost regressor framework and designed a test-train-validation strategy to account for these issues **(Figure 7A)**, exploring many features representing conformation, stability, inhibitor-mutation distance, and inhibitor chemical information **(Figure 7 - Figure Supplement 1)**. To avoid overfitting, we introduced several constraints on the monotonicity and the precision of certain features. The final model uses a subset of the features we tested and improves the performance from 0.28 to 0.37 **(Figure 7B,C)**. The model primarily improves the correlation by shifting the distribution of predicted fitness values to center around drug sensitivity, reflecting the pressures that are not accounted for by ESM-1b **(Figure 7D)**. Nonetheless, many resistant mutations are correctly predicted by the new model.

To examine whether the model could help interpret the mechanisms of specific mutations, we examined several cases with notable improved predictions as the model increased in complexity **(Figure 7E,F)**. For some mutations, as in Y1230D, we observe a gradual improvement in prediction for each set of features, suggesting that resistance relies on multiple factors. For other mutations, such as N1167K, we see a single set of features driving the improvement, which suggests much more dominant driving forces. Lastly, in other mutations, like G1290D, the models trained with different features can over or under predict the true value, demonstrating the value of combining features together. The reliance on simple features helps identify some of the major factors in drug resistance and sensitization such as distance to the inhibitor and active/inactive conformation; however, improved feature engineering and coverage of both sequence and chemical space will likely be needed to create a more interpretable model.

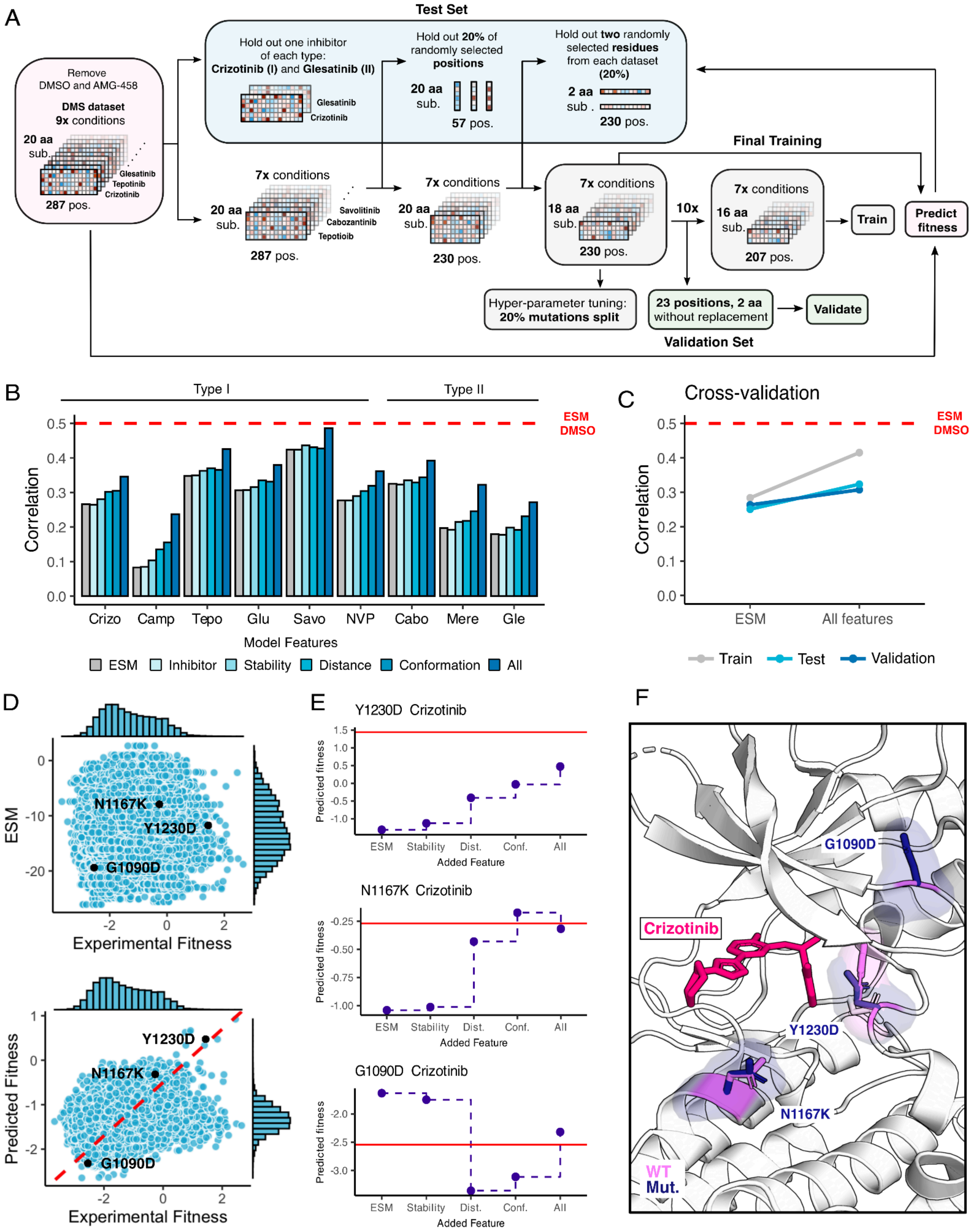

A
Test Set
Remove DMSO and AMG-458
DMS dataset 9x conditions
20 aa sub.
Glesatinib
Tepotinib
Crizotinib
287 pos.
Hold out one inhibitor of each type: Crizotinib (I) and Glesatinib (II)
Glesatinib
Crizotinib
Hold out 20% of randomly selected positions
20 aa sub.
57 pos.
Hold out two randomly selected residues from each dataset (20%)
2 aa sub .
230 pos.
7x conditions
20 aa sub.
Savolitinib
Cabozantinib
Tepotioib
287 pos.
7x conditions
20 aa sub.
230 pos.
Final Training
7x conditions
18 aa sub.
230 pos.
10x
7x conditions
16 aa sub.
207 pos.
Train
Predict fitness
Hyper-parameter tuning: 20% mutations split
23 positions, 2 aa without replacement
Validate
Validation Set
B
Type I
Type II
ESM DMSO
Correlation
0.0
0.1
0.2
0.3
0.4
0.5
Crizo
Camp
Tepo
Glu
Savo
NVP
Cabo
Mere
Gle
Model Features
ESM
Inhibitor
Stability
Distance
Conformation
All
C
Cross-validation
ESM DMSO
Correlation
0.0
0.1
0.2
0.3
0.4
0.5
ESM
All features
Train
Test
Validation
D
ESM
0
-10
-20
N1167K
Y1230D
G1090D
-2
0
2
Experimental Fitness
Predicted Fitness
1
0
-1
-2
Y1230D
N1167K
G1090D
-2
0
2
Experimental Fitness
E
Y1230D Crizotinib
Predicted fitness
1.5
1.0
0.5
0.0
-0.5
-1.0
ESM
Stability
Dist.
Conf.
All
Added Feature
N1167K Crizotinib
Predicted fitness
-0.25
-0.50
-0.75
-1.00
ESM
Stability
Dist.
Conf.
All
Added Feature
G1090D Crizotinib
Predicted fitness
-2.0
-2.5
-3.0
ESM
Stability
Dist.
Conf.
All
Added Feature
F
G1090D
Crizotinib
Y1230D
N1167K
WT
Mut.

Figure 7. Inhibitor-bound variant fitness predicted from a machine learning model trained on the MET DMS dataset. (A) Model architecture outlining the information flow and inputs for model training, validation, fitness predictions, and prediction tests. (B) Improvement in correlation between experimental and predicted fitness for each inhibitor with usage of different kinds of features. (C) Cross-validation trends between the baseline ESM model and the model with all features incorporated. (D) Scatter plots of predictions versus experimental fitness scores of the baseline ESM model (top) compared to the model with all features (bottom), with a dashed cross-graph line in red displayed. (E) Residue-level analysis of feature significance in fitness predictions (ESM, stability, distance, conformation, all features). The Rosace experimental score is shown as a red line. (F) Residues with improved predictions mapped on a crizotinib-bound MET kinase domain (PDB 2WGJ). Predicted resistance mutations (dark purple) modeled relative to the wild-type residue (pink).

Discussion

Tyrosine kinase inhibitors have revolutionized the treatment of many diseases, but the development of resistance creates a significant challenge for long term efficacy. Many strategies, including sequential dosing **(Attwood et al., 2021; Recondo et al., 2020)**, are being explored to overcome resistance. Our DMS of the MET receptor tyrosine kinase domain, performed against a panel of varying inhibitors offers a framework for experimentally identifying resistance and sensitizing mutations in an activated kinase context for different inhibitors. By massively screening the effect of a nearly comprehensive library of amino acid mutations in the MET kinase domain against 11 inhibitors, some generalizable patterns emerged. In concordance with the binding mode of both type I and II inhibitors, residues that commonly confer resistance, or act as “hotspots,” were mapped to previously reported sites like D1228, Y1230, M1211, G1163, and novel sites like I1084, L1140, L1142, T1261, and L1272. Annotation of hotspots also offers an opportunity to inform inhibitor selection based on likelihood of cross-inhibitor resistance **(Figure 5)**. For instance, I1084 is a hotspot for type I and II inhibitors within our study that displayed wild-type sensitivity to the type I½ inhibitor screened **(Figure 5)**. Understanding positions with high resistance frequencies that are distal from the ATP-binding site also offers a design opportunity for allosteric inhibitors that can target cancer-associated and resistance-associated regions within the N- and C-lobe **(Mingione et al., 2023)**.

Nevertheless, similar to its ability in identifying resistance for inhibitors, our parallel DMS also demonstrated the ability to detect non-selective drugs, with the example of tivantinib. Despite being a proposed MET-selective inhibitor, like several others, tivantinib failed clinical trials, and follow-up studies suggested cytotoxicity and off-target binding as the culprit **(Michieli and Di Nicolantonio, 2013; Basilico et al., 2013; Katayama et al., 2013; Fujino et al., 2019)** - a scenario that is not uncommon to antitumor drugs that do not advance to the clinic **(Lin et al., 2019)**. To this effect, the ability of DMS to differentiate between selective compounds provides a unique prospect for developing inverse structure−activity relationships, whereby varying protein sequence both inhibitor specificity and resistance can be learned.

Reported cancer mutations in databases such as OncoKB or cBioPortal are useful for patient data and cancer type reporting **(Suehnholz et al., 2024; Chakravarty et al., 2017; Cerami et al., 2012)**. A recent analysis of these databases aided in annotation of mutations observed within patient populations **(Pecci et al., 2024)**. This study used pre-clinical models to examine a subset of these mutations and identified sensitivities to multiple inhibitors and confirmed clinical responses of two rare driver mutations (H1094Y and F1200I) to elzovantinib, a type Ib inhibitor **(Pecci et al., 2024)**. Their results are consistent with the predictions of our DMS, illustrating the potential value of having a broad dictionary of inhibitor sensitivity and resistance patterns.

Finally, a significant challenge of inhibitor screening is the considerable time and cost involved, even at high-throughput. While docking has accelerated the prioritization of compounds for protein targeting and screening *in silico* **(Sadybekov and Katritch, 2023)**, prediction of drug resistance is of high interest in informing iterative drug design. Screening for resistance in the early stages of drug design is particularly useful for obtaining inhibitors that can be effective in the long-term by optimizing protein-inhibitor interactions in the wildtype and functionally silent mutant context **(Pisa and Kapoor, 2020)**. While base-editor approaches can rapidly screen for inhibitor resistance mutations within full-length, endogenous genes, undersampling of rare variants due to lower coverage is a significant caveat **(Dorighi et al., 2024)**, compared to DMS where nearly full coverage is achieved and controlled. A full landscape of mutational effects can help to predict drug response and guide small molecule design to counteract acquired resistance. The ideas motivating our ML-model, which combines protein language models and biophysical/chemical features, to novel inhibitors could eventually be used to profile resistance and sensitivity for novel and unscreened small molecules, greatly extending the scale of kinase inhibitor repositioning for second-line therapies.

Materials and methods

Mammalian cell culturing

Ba/F3 cells (DSMZ) were maintained and passaged in 90% RPMI (Gibco), 10% HI-FBS (Gibco), 1% penicillin/streptomycin (Gibco), and 10ng/ml IL-3 (Fisher), and incubated at 37°C with 5% CO2. Cells were passaged at or below 1.0e6 cells/ml to avoid acquired IL-3 resistance, and regularly checked for IL-3 dependence by performing 3x PBS (Gibco) washes and outgrowth in the absence of IL-3.

Plat-E cells stably expressing retroviral envelope and packaging plasmids were originally gifted by Dr. Wendell Lim, and maintained in 90% DMEM, HEPES (Gibco), 10% HI-FBS (Gibco), 1% penicillin/streptomycin (Gibco), 10ug/ml blasticidin, 1ug/ml puromycin. Cells were cultured at 37°C with 5% CO2 and maintained under blasticidin and puromycin antibiotic pressure unless being transfected.

Dose response and IC50 determination of inhibitors

Unless otherwise stated, all inhibitors used in this study were purchased from SelleckChem.

Ba/F3 cells stably expressing TPR-MET and TPR-METΔEx14 were washed with DPBS (Gibco) 3x times to remove IL-3, puromycin, penicillin, and streptomycin. Cells were resuspended in 90% RPMI and 10% FBS, and were seeded in the wells of a 96-well, round-bottom plate at a density of 2.5e4 cells/ml in 200ul. Cells were incubated for 24hr to allow kinase-driven signaling. The next day, inhibitors were added to triplicate rows of cells at a concentration range of 0 to 10uM (2-fold dilutions), and allowed to incubate for 72 hr post TKI addition. CellTiter-Glo reagent (Promega) was mixed at a 1:1 ratio with cells to lyse and provide a luminescence readout, which was measured on a Veritas luminometer. Cell numbers were determined from a Ba/F3 cell and ATP standard curve generated according to the manufacturer's instructions. Dose response curves were fitted using GraphPad Prism with the log(inhibitor) vs. response, variable slope function. Data are presented as cell viability normalized to the fold change from the TKI free control.

MET kinase domain variant library generation, cloning, and library introduction into Ba/F3

In this study, we repurposed cell lines transduced with TPR-MET and TPR-METΔEx14 kinase domain variant libraries, previously reported in Estevam et al., 2024. All libraries were generated, transfected, and tested in parallel.

In short, the MET kinase domain sequence used in this study spans amino acid positions 1059-1345, which includes the full kinase domain (aa 1071-1345) and a small region of the juxtamembrane (aa 1059-1070). The variant DNA library was synthesized by Twist Bioscience, containing one mammalian high-usage codon per amino acid. A “fill-in” library was generated to introduce an early stop control codon every 11 amino acids evenly spaced across the sequence. In addition, mutations at positions with failed synthesis (positions 1194 and 1278) were generated and added at equimolar concentration into the variant library. The kinase domain variant library was introduced into two different cloning backbones, one carrying the TPR-fusion sequence with the wild-type juxtamembrane sequence (aa 963-1058), wild-type C-terminal tail (aa 1346-1390), and IRES-EGFP (pUC19_kozak-TPR-METΔEx14-IRES-EGFP) and the other carrying the TPR-fusion sequence with an exon 14 skipped juxtamembrane sequence (aa 1010-1058), wild-type C-terminal tail (aa 1346-1390), and IRES-mCherry (pUC19_kozak-TPR-MET-IRES-mCherry). The libraries were transformed in MegaX 10 beta cells (Invitrogen), propagated in 50mL LB and Carbinacillin at 37°C to an OD of 0.5, and then midiprepped (Zymo). Library coverage was determined by colony count of serial dilutions from the recovery plated at varying dilutions (1:100, 1:1k, 1:10k, 1:100k, 1:1M).

The full TPR-METΔEx14-IRES-EGFP and TPR-MET-IRES-mCherry variant libraries were then shuttled into the mammalian retroviral MSCV backbone (addgene) through restriction enzyme digest with MluI-HF (NEB) and MfeI-HF (NEB), then ligated into the empty backbone with T4 ligase (NEB). Ligations were DNA cleaned (Zymo), electroporated into ElectroMAX Stbl4 Competent Cells (Thermo Fisher), plated on LB-agar bioassay plates with Carbenicillin, incubated at 37°C, then colonies were scraped into 50mL LB and midi-prepped for transfections (Zymo).

Variant libraries were transfected into Plat-E cells for retroviral packaging using Lipofectamine3000 (Invitrogen) following the manufacturer's for a T-175 scale, and using a total of 46 µg DNA. 48 hr post-transfection, the viral supernatant was harvested, passed through a 0.45 µm sterile filter, then concentrated with Retro-X concentrator (TakaraBio) using a 1:4 ratio of concentrator to supernatant. The concentrated virus was titered in Ba/F3 to determine the proper volume for a transduction MOI of 0.1-0.3. The viral titer was calculated from the percent of fluorescent cells and viral dilution. To generate the DMS transduced cell lines, 6 million cells were spinfected at an MOI of 0.1, in triplicate. Then infected cells were selected with 1µg/ml puromycin in 4 days, with fluorescence and cell counts tracked each day.

DMS time point selection and sample preparation

All screening conditions were performed and handled in parallel for TPR-MET and TPR-METΔEx14 libraries across all independent conditions and biological replicates.

For each biological replicate, a stock of 4.0e6 cells transduced with TPR-MET and TPR-METΔEx14 kinase domain variants was thawed and expanded for 48 hrs in the presence of IL-3 and puromycin to prevent pre-TKI selection to reach a density for screen seeding. Each batch of cells were grown to a density of 72 million cells to be split into 12 dishes (15cm) for each selection condition. Cells were first washed with DPBS (Gibco) 3 times to remove IL-3 and antibiotics. Cells were resuspended in 90% RPMI and 10% FBS, counted, and split across 12 dishes (15cm) at a density of 6 million cells in 30mL. A total of 6 million cells from each replicate was harvested and pelleted at 250xg to serve as the “time point 0” pre-selection sample (T0).

To begin selection of each replicate for each library, DMSO was added to the control plate (0.01% final) while the appropriate IC50 concentration of inhibitor was added to each respective plate (independent pool of cells). Three time points post T0 were collected for each library replicate and inhibitor condition for a total of 4 time points (T0, T1, T2, T3). Time points were harvested every two doublings (~72hr) across 12 days; 6 million cells were harvested for each condition and pelleted at 250xg for 5min; 2.0e5 cells/ml were split at every time point and maintained either in DMSO or TKI at the appropriate concentration to maintain cellular growth rates under inhibitor selection.

The gDNA of each time point sample was isolated with the TakaraBio NucleoSpin Blood QuickPure kit the same day the cells were harvested. gDNA was eluted in 50µl of elution buffer provided by the kit, using the high concentration and high yield elution manufacturer's protocol. Immediately after gDNA was isolated, 5µg of gDNA was used for PCR amplification of the target MET KD gene to achieve the proper variant coverage. A 150µl PCR master mix was prepared for each sample using the TakaraBio PrimeStar GXL system according to the following recipe: 30µl 5X PrimeStar GXL buffer, 4.5µl 10µM forward primer (0.3uM final), 4.5µl 10µM reverse primer (0.3uM final), 5µg gDNA, 12µl 10mM dNTPs (2.5mM each NTP), 6µl GXL polymerase, nuclease free water to a final reaction volume of 150uL. The PCR master mix for each sample was split into three PCR tubes with 50µl volumes for each condition and amplified with the following thermocycler parameters: initial denaturation at 98°C for 30 s, followed by 24x cycles of denaturation at 98°C for 10 s, annealing at 60°C for 15 s, extension at 68°C for 14 s, and a final extension at 68°C for 1 min.

PCR samples were stored at -20°C until all time points and replicates were harvested and amplified, so as to prepare all final samples for NGS together with the same handling and sequence them in the same pool to prevent sequencing bias.

Library preparation and Next-generation sequencing

After all time points were selected, harvested, and PCR amplified, the target gene amplicon was isolated from gDNA by gel purification (Zymo), for a total of 222 samples. The entire 150µl PCR reaction for each sample was mixed with 1X NEB Purple Loading Dye (6X stock) and run on a 0.8% agarose, 1X TBE gel, at 100 mA until there was clear ladder separation and distinct amplicon bands. The target amplicons were gel excised and purified with the Zymo Gel DNA Recovery kit. To remove excess agarose contamination, each sample was then further cleaned using the Zymo DNA Clean and Concentrator-5 kit and eluted in nuclease free water. Amplicon DNA concentrations were then determined by Qubit dsDNA HS assay (Invitrogen).

Libraries were then prepared for deep sequencing using the Nextera XT DNA Library Prep kit in a 96-well plate format (Illumina). Manufacturer's instructions were followed for each step: tagmentation, indexing and amplification, and clean up. Libraries were indexed using the IDT for Nextera Unique Dual Indexes Set A,B and C (Illumina). Then, indexed libraries were quantified using the Agilent TapeStation with HS D5000 screen tape (Agilent) and reagents (Agilent). DNA concentrations were further confirmed with a Qubit dsDNA HS assay (Invitrogen). All samples were manually normalized and pooled at 10nM (MET and METΔEx14 in the same pool). The library was then paired-end sequenced (SP300) on two lanes of a NovaSeq6000.

MET kinase domain variant analysis and scoring

Enrich2 Scoring

Our approach followed the one used for our initial MET DMS experiments [cite Met 1]. Sequencing files were obtained from the sequencing core as demultiplexed fastq.gz files. The reads were first filtered for contamination and adapters using BBDuk, then the paired reads were error-corrected and merged with BBMerge and mapped to the reference sequence using BBMap (all from BBTools) [Bushnell, B. (2014). BBTools software package.]. Read consequences were determined and counted using the AnalyzeSaturationMutagenesis tool in GATK v4 **(van der Auwera and O'Connor, 2020)**. This is further processed by use of a script to filter out any variants that are not expected to be in the library (i.e., variants due to errors in sequencing, amplification, etc). The final processed count files were then analyzed with Enrich2, using weighted least squares and normalizing to wildtype sequences **(SRA 28784151)**.

Rosace scoring

We used Rosace to analyze experiments of different conditions (DMSO or inhibitors) independently. In order to make the scores more comparable and interpretable across conditions, we modified the original Rosace software so that the output scores reflect the scale of cell doubling rate between every contiguous time point. For example, in an ideal experiment, if the wild-type cell doubling rate is 2 and its score is 0 by wild-type normalization, a score of -2 means that the cells are not growing (2^(2-2)) and a score of 1 means that the cells are doubling three times (increasing to 2^(2+1) times the original count) between every contiguous time point.

The input to Rosace is the filtered count files provided by the AnalyzeSaturationMutagenesis tool described in the above section. From here, we filtered variants by the mean count (>= 4) and the proportion of 0 count across replicates and time points (<= 10/12) and before inhibitor selection at T0 (<= 2/3). Second, we normalized the counts using wild-type normalization with log2 transformation rather than the default natural log transformation to maintain the doubling rate scale. Finally, the normalized count is regressed on time intervals (t = {1, 2, 3, 4}) instead of the entire time span (t = {1/4, 2/4, 3/4, 4/4}) so that the resulting score reflects the growth rate between every contiguous time point.

Statistical filtering and resistance classification

In mathematical terms, we define the raw Rosace fitness scores of a mutation in DMSO as $\beta_{v,DMSO}$ and in a specific inhibitor condition as $\beta_{v,inh}$. The scores of wildtype variants were normalized to 0, and we denote them as $\beta_{wt,DMSO} = 0$ and $\beta_{wt,inh} = 0$. Growth rate of wild-type cells under different inhibitor selections were controlled to be identical (two doublings between every time point), so even though raw Rosace scores are computed independently per condition, $\beta_{v,inh}$ are directly comparable between inhibitor conditions.

Within one condition (DMSO or inhibitor), according to convention, we call variants with $\beta_{v,inh} \gg 0$ or $\beta_{v,DMSO} \gg 0$ "gain-of-function" and $\beta_{v,inh} \ll 0$ or $\beta_{v,DMSO} \ll 0$ "loss-of-function". With scores from multiple conditions, we presented three types of filtering strategies and produced the following classification: inhibitor-specific "resistance mutation", inhibitor-specific "resistance position", and "loss-of-function" and "gain-of-function" mutation in the context of growth rate differential with and without an inhibitor.

We stress the different interpretations of "gain-of-function" and "loss-of-function" labels. Within one condition, this label is a general term to describe whether the function of protein is perturbed by the mutation. In contrast, the latter describes the difference with and without a given inhibitor, canceling effects of folding, expression, and stability and targeting only the inhibitor sensitivity function of the protein.

A "resistance mutation" is specific to a certain inhibitor, and it satisfies the chained inequality $\beta_{v,inh} \gg \beta_{wt,inh} = \beta_{wt,DMSO} \geq \beta_{v,DMSO}$. The first inequality specifies that the growth rate of a resistant mutation needs to be much larger than that of the wild-type in the presence of the inhibitor, and we used the one-sided statistical test $\beta_{v,inh} > 0.5$ with the test statistics cutoff 0.1. The second inequality specifies that in DMSO, the growth rate of that mutation is equal to or lower than that of the wild-type, ensuring that the resistance behavior we see is specific to that inhibitor, not that it grows faster under every condition, and thus we used the effect size cutoff $\beta_{v,DMSO} \leq 0$.

A "resistance position" is a position that contains at least one "resistance mutation" to a certain inhibitor.

In the context of growth rate differential with and without an inhibitor, a mutation is "gain-of-function" if it has a higher growth rate in the presence of the inhibitor than in its absence, which is one feature of "resistance mutation". It is "loss-of-function" if it grows faster in the absence of the inhibitor. To label the mutations accordingly, we first computed a recentered Rosace score for each mutation under inhibitor selection $\gamma_{v,inh} = \beta_{v,inh} - \beta_{v,DMSO}$, and define "gain-of-function" $\gamma_{v,inh} > 0.75$ and "loss-of-function" $\gamma_{v,inh} < 0$ in the differential sensitivity analysis.

Machine learning modeling

Feature selection for the machine learning model

Interpretable features of the MET sequence variants and inhibitors were carefully chosen to be incrementally added to a model. To extract structural features from inhibitor bound mutant complexes, we used Umol to predict the structures of all the MET kinase variants bound to each of the inhibitors **(Bryant et al. 2023)**. The input to Umol is the MET kinase variant sequence, SMILES string of the inhibitor and list of residues lining the putative binding pocket. The predicted complexes (MET kinase bound to inhibitor) were relaxed using OpenMM **(Eastman et al. 2013)**. To ensure the inhibitor in the predicted structures are in the same pose as compared to reference structures, we tethered the predicted inhibitor structure to the reference pose using a modified version of the script available in https://github.com/Discngine/rdkit_tethered_minimization. The reference pose for crizotinib, NVP-BVU972, Merestenib and Savolitinib were taken from the corresponding crystal structures in the PDB - 2WGJ, 3QTI, 4EEV and 6SDE, while the reference pose for cabozantinib, capmantinib, glumetinib and glesatinib analog were taken from the structures docked using Autodock Vina (see Kinase domain structural analysis). Following this, the tethered inhibitors were redocked back to the predicted variant structures using Autodock Vina **(Eberhardt et al. 2021)**. We also extracted features from wild-type MET kinase structures. The features could be broadly classified into four categories: inhibitor, stability, distance, conformation and inhibitor binding. Apart from these, ESM Log Likelihood Ratio was used as a feature in all models that we trained. Each of the feature categories that we explored and the rationale behind choosing them are explained below:

(1) ESM Log Likelihood Ratio (ESM LLR)

ESM1b is an unsupervised protein language model trained on a large set of protein sequences from UniProt that has successfully learned protein fitness patterns **(Rives et al, 2021; Lin et al., 2023)**. By including a mask token at a given position in the sequence, the log-likelihoods of all amino acid substitutions can be extracted from the model. The ratio between ESM1b log-likelihoods for the mutant and wildtype amino acids provides a score that indicates the fitness of each variant in the mutational scan, with log-likelihood ratios having precedent as a variant predictor **(Rives et al, 2021; Lin et al., 2023)**. The predictions used here were obtained using esm-variants webserver (https://huggingface.co/spaces/ntranoslab/esm_variants) **(Brandes et al., 2023)**.

(2) Inhibitor features:

- **Inhibitor molecular weight**: We calculated the molecular weight of each inhibitor as a feature.
- **Ligand RMSD**: We structurally superposed the predicted variant structure onto the corresponding wildtype structure and calculated the RMSD between the predicted, re-docked inhibitor and the reference inhibitor structure **(Figure 7 - figure supplement 1J)**

(3) Stability features:

- **ΔΔΔG and ΔΔG**: Because inhibitor types are largely distinguished based on binding configuration, we reasoned that the difference in stability contributed by each mutation between given binding states (e.g. Type I bound state vs. a Type II bound state) could contribute to the success of the predictor. To compute the stability difference, we used structural representatives for type-I bound (2WGJ) and type II bound (4EEV) MET kinase and calculated the change in free energy (ΔΔG) of every possible mutation at every position using ThermoMPNN **(Dieckhaus et al., 2023).** The difference in ΔΔG between type-I bound and type-II bound structures (ΔΔΔG) for every variant was added as a feature to the XGBoost model to capture the difference in stabilization from the mutation in the Type I or Type II bound state (**Figure 7 - figure supplement 1C**). We also used the predicted ΔΔG score of the corresponding inhibitor type-bound structure directly as a feature. For instance, if the input data corresponds to a mutation to Alanine at position 1065 in the presence of glumetinib (a type I inhibitor), the difference between ΔΔG predicted for the 1065A variant for the type-I bound (2WGJ) and for type II bound (4EEV) structure is used as a feature (ΔΔΔG). The ΔΔG predicted for the 1065A variant for the type-I bound (2WGJ) structure is also used as a feature.

(4) Distance features:

- **Residue to ATP distance**: Proximity to the ATP-binding site indicates the ability of the given residue to influence inhibitor binding given that Type I and Type II inhibitors are ATP competitive. To include this feature, the distance between C-alpha residue atoms and the centroid of bound ATP in a representative structure (PDB 3DKC) was calculated and the distance corresponding to each position was added as a feature **(Figure 7 - figure supplement 1D)**.
- **Inhibitor distance**: This is the shortest distance between the inhibitor and mutated residue in the predicted variant-inhibitor complexes. **(Figure 7 - figure supplement 1G)**.

(5) Conformational features:

- **MET crystal structure RMSF:** The extent of flexibility at the mutation position could be significantly affected by the mutation, which in turn can affect the function of the variant. To account for this, we utilized the structural information abundantly available for MET kinases in PDB. We structurally aligned all crystal structures of human MET kinases with resolution better than 3Å (81 structures) using mTM-align **(Dong et al., 2018)** and calculated the Root Mean Squared Fluctuation at every residue position using Prody **(Zhang et al., 2021) (Figure 7 - figure supplement 1F).**
- **Residue RMSD:** We structurally superposed the predicted variant structure onto the corresponding wildtype structure and calculated the RMSD between the mutant and wildtype residue at the mutation position **(Figure 7 - figure supplement 1I)**

(6) Inhibitor Binding features:

- **RF-Score**: To quantify the binding strength between the inhibitor and the variant protein structure, we calculated the RF-score, which is a random forest-based approach to predict protein-ligand binding affinity **(Wójcikowski et al., 2017)**
- **Pocket volume**, **hydrophobicity score** and **polarity score:** Changes to the binding pocket in terms of volume and hydrophobicity due to mutations could affect the interaction and binding between the inhibitor and variant. These effects were brought in as features into the model by calculating the binding pocket volume, hydrophobicity score and polarity score of the binding pocket using fpocket **(Le Guilloux et al., 2009) (Figure 7 - figure supplement 1H)**.

This category of features are not part of the best performing model shown in Figure 7.

Apart from these categories, we calculated the difference in volume between the wildtype and mutated residue at a given position and added it as a feature (∆**Volume**) since residue volume changes upon mutation could contribute to steric hindrance **(Figure 7 - figure supplement 1E)**. This feature is also not part of the best performing model.

This led to a total of 14 interpretable features to evaluate our models on. We trained and tested a total of 8192 models by considering all possible numbers and combinations of these features (keeping ESM LLR as a constant feature in all models). The hyperparameter tuning, cross-validation, training and testing of each of these models are described in detail below.

Training and Selecting the Predictive Model

An XGBoost regressor model, which is a gradient boosting method based on decision trees as the base learner **(Chen and Guestrin, 2016)**, was used to predict DMS fitness scores in presence of inhibitors. Given the relatively small dataset we are using here, the models are prone to overfitting. Hence, we used monotonic constraints on features that had a monotonic relationship with the experimental fitness scores. ESM LLR score and ∆∆G have a positive and negative correlation with the experimental fitness scores respectively **(Figure 7 - figure supplement 1A)**. Therefore, ESM LLR was constrained positively and ∆∆G was constrained negatively by assigning 1 and -1 respectively to the ‘monotone_constraints’ parameter in Python XGBoost. This ensures that the monotonic relationship between the input feature and the target value is maintained during predictions.To further prevent overfitting, we binned the values of the 12 remaining into four or five bins and assigned the median of the bin as their value. The bins were chosen such that one or two bins would contain the majority of feature values. The distribution of these twelve features are shown in **Figure 7 - figure supplement 1B**. The bins of each feature are shown as red dashed lines on the histograms. Model performance was evaluated using Pearson’s R and mean squared error (MSE).

Experimental fitness scores of MET variants in the presence of DMSO and AMG458 were ignored in model training and testing since having just one set of data for a type I ½ inhibitor and DMSO leads to learning by simply memorizing the inhibitor type, without generalizability. The remaining dataset was split into training and test sets to further avoid overfitting **(Figure 7A)**. The following data points were held out for testing - (a) all mutations in the presence of one type I (crizotinib) and one type II (glesatinib analog) inhibitor, (b) 20% of randomly chosen positions (columns) and (c) all mutations in two randomly selected amino acids (rows) (e.g. all mutations to Phe, Ser). After splitting the dataset into train and test sets, the train set was used for XGBoost hyperparameter tuning and cross-validation. For tuning the hyperparameters of each of the XGBoost models, we held out 20% of randomly sampled data points in the training set and used the remaining 80% data for Bayesian hyperparameter optimization of the models with Optuna **(Akiba et al., 2019)**, with an objective to minimize the mean squared error between the fitness predictions on 20% held out split and the corresponding experimental fitness scores. The following hyperparameters were sampled and tuned: type of booster (booster - gbtree or dart), maximum tree depth (max_depth), number of trees (n_estimators), learning rate (eta), minimum leaf split loss (gamma), subsample ratio of columns when constructing each tree (colsample_bytree), L1 and L2 regularization terms (alpha and beta) and tree growth policy (grow_policy - depthwise or lossguide). After identifying the best combination of hyperparameters for each of the models, we performed 10-fold cross validation (with re-sampling) of the models on the full training set. The training set consists of data points corresponding to 230 positions and 18 amino acids. We split these into 10 parts such that each part corresponds to data from 23 positions and 2 amino acids. Then, at each of 10 iterations of cross-validation, models were trained on 9 of 10 parts (207 positions and 16 amino acids) and evaluated on the 1 held out part (23 positions and 2 amino acids). Through this protocol we ensure that we evaluate performance of the models with different subsets of positions and amino acids. The average Pearson correlation and mean squared error of the models from these 10 iterations were calculated and the best performing model out of 8192 models was chosen as the one with the highest cross-validation correlation. The final XGBoost models were obtained by training on the full training set and also used to obtain the fitness score predictions for the validation and test sets. These predictions were used to calculate the inhibitor-wise correlations shown in Figure 7B.

Kinase domain structural analysis

Unless otherwise stated, all structural analysis was performed on PyMOL. Structural mapping incorporated tools from the Bio3D bioinformatics package in R **(Grant et al., 2006).** Inhibitors that lacked an experimental crystal structure were docked into a representative type I (2WGJ) or type II (4EEV) structure with AutoDock Vina **(Eberhardt et al., 2021)**. Existing ligands in both the structures were removed *in silico* and the proteins prepared for docking using AutoDockTools by adding polar hydrogens and Kollman charges. The inhibitors were also prepared using AutoDockTools by adding polar hydrogens and charges and identifying rotatable torsions. A grid box which dictates the search space for the docking tool was defined approximately around the region where the existing ligands in 2WGJ and 4EEV were bound. The energy range and exhaustiveness of docking was set to 3 and 8 respectively. AutoDock Vina was made to output 5 modes for each ligand. Capmatenib and glumetinib (type I inhibitors) were docked on to 2WGJ and glesatinib analog and cabozantinib (type II inhibitors) were docked on to 4EEV.

Acknowledgements

Sequencing was performed at the UCSF CAT, supported by UCSF PBBR, RRP IMIA, and NIH 1S10OD028511-01 grants. This work was supported by NIH CA239604 to EAC, NJ, JSF; HHMI Hanna Gray Fellowship and UCSF QBI Fellow program to WCM; NIH R01LM013434 to JAC; NIH GM145238 and the UCSF Program for Breakthrough Biomedical Research, funded in part by the Sandler Foundation, to JSF.

Author contributions

Conceptualization JSF, NJ, EAC; Methodology, GOE, EML, JR, CBM; Formal analysis, GOE; Writing – Original Draft, GOE and JSF; Writing – Editing, GOE, JSF, EML, NJ, EAC; Validation, GOE, EML; Machine learning, AR, KC; Funding Acquisition, JSF, NJ, EAC; Resources, JSF, NJ, EA, WCM, HP; Supervision, JSF, NJ, EAC, WMC, EML, HP, JAC.

Competing Interests

JSF is a consultant for, has equity in, and receives research support from Relay Therapeutics and is a consultant for Octant Bio. N.J. is a founder of Rezo Therapeutics and a shareholder of Rezo Therapeutics, Sudo Therapeutics, and type6 Therapeutics. N.J. is a SAB member of Sudo Therapeutics, type6

Therapeutic and NIBR Oncology. The Jura laboratory has received sponsored research support from Genentech, Rezo Therapeutics and type6 Therapeutics. E.A.C. is a consultant at IHP Therapeutics, Valar Labs, Tatara Therapeutics and Pear Diagnostics, reports receiving commercial research grants from Pfizer, and has stock ownership in Tatara Therapeutics, HDT Bio, Clara Health, Aqtual, and Guardant Health.

Code and data availability

The sequencing data has been deposited at the NCBI SRA (28784151). Original data files, analysis, and source code is available at https://github.com/fraser-lab/MET_kinase_Inhibitor_DMS

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. https://doi.org/10.48550/arXiv.1907.10902

An, L., Wang, Y., Wu, G., Wang, Z., Shi, Z., Liu, C., Wang, C., Yi, M., Niu, C., Duan, S., Li, X., Tang, W., Wu, K., Chen, S., & Xu, H. (2023). Defining the sensitivity landscape of EGFR variants to tyrosine kinase inhibitors. *Translational Research*, *255*, 14–25. https://doi.org/10.1016/j.trsl.2022.11.002

Arter, C., Trask, L., Ward, S., Yeoh, S., & Bayliss, R. (2022). Structural features of the protein kinase domain and targeted binding by small-molecule inhibitors. *Journal of Biological Chemistry*, *298*(8), 102247. https://doi.org/10.1016/j.jbc.2022.102247

Attwood, M. M., Fabbro, D., Sokolov, A. V., Knapp, S., & Schiöth, H. B. (2021). Trends in kinase drug discovery: Targets, indications and inhibitor design. *Nature Reviews Drug Discovery*, *20*(11), Article 11. https://doi.org/10.1038/s41573-021-00252-y

Auwera, G. van der, & O’Connor, B. D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. O’Reilly Media, Incorporated.

Bahcall, M., Paweletz, C. P., Kuang, Y., Taus, L. J., Sim, T., Kim, N. D., Dholakia, K. H., Lau, C. J., Gokhale, P. C., Chopade, P. R., Hong, F., Wei, Z., Köhler, J., Kirschmeier, P. T., Guo, J., Guo, S., Wang, S., & Jänne, P. A. (2022). Combination of Type I and Type II MET Tyrosine Kinase Inhibitors as Therapeutic Approach to Prevent Resistance. *Molecular Cancer Therapeutics*, *21*(2), 322–335. https://doi.org/10.1158/1535-7163.MCT-21-0344

Bahcall, M., Sim, T., Paweletz, C. P., Patel, J. D., Alden, R. S., Kuang, Y., Sacher, A. G., Kim, N. D., Lydon, C. A., Awad, M. M., Jaklitsch, M. T., Sholl, L. M., Jänne, P. A., & Oxnard, G. R. (2016). Acquired *MET* D1228V Mutation and Resistance to MET Inhibition in Lung Cancer. *Cancer Discovery*, *6*(12), 1334–1341. https://doi.org/10.1158/2159-8290.CD-16-0686

Basilico, C., Pennacchietti, S., Vigna, E., Chiriaco, C., Arena, S., Bardelli, A., Valdembri, D., Serini, G., & Michieli, P. (2013). Tivantinib (ARQ197) Displays Cytotoxic Activity That Is Independent of Its Ability to Bind MET. *Clinical Cancer Research*, *19*(9), 2381–2392. https://doi.org/10.1158/1078-0432.CCR-12-3459

Brandes, N., Goldman, G., Wang, C. H., Ye, C. J., & Ntranos, V. (2023). Genome-wide prediction of disease variant effects with a deep protein language model. *Nature Genetics*, *55*(9), Article 9. https://doi.org/10.1038/s41588-023-01465-0

Brazel, D., Zhang, S., & Nagasaka, M. (2022). Spotlight on Tepotinib and Capmatinib for Non-Small Cell Lung Cancer with MET Exon 14 Skipping Mutation. *Lung Cancer: Targets and Therapy*, *13*, 33–45. https://doi.org/10.2147/LCTT.S360574

Brenan, L., Andreev, A., Cohen, O., Pantel, S., Kamburov, A., Cacchiarelli, D., Persky, N. S., Zhu, C., Bagul, M., Goetz, E. M., Burgin, A. B., Garraway, L. A., Getz, G., Mikkelsen, T. S., Piccioni, F., Root, D. E., & Johannessen, C. M. (2016). Phenotypic Characterization of a Comprehensive Set of MAPK1/ERK2 Missense Mutants. *Cell Reports*, *17*(4), 1171–1183. https://doi.org/10.1016/j.celrep.2016.09.061

Bryant, P., Kelkar, A., Guljas, A., Clementi, C., & Noé, F. (2023). *Structure prediction of protein-ligand complexes from sequence information with Umol* (p. 2023.11.03.565471). bioRxiv. https://doi.org/10.1101/2023.11.03.565471

Cai, B., Li, X., Huang, X., Ma, T., Qu, B., Yu, W., Yang, W., Zhang, P., Chen, J., & Liu, F. (2021). Case Report: Sequential Combination Targeted Therapy With Type I and II MET Inhibitors in a Metastatic EGFR-Mutated, MET-Amplified NSCLC Patient With Acquired MET Y1230H Mutation. *Frontiers in Oncology*, *11*. https://www.frontiersin.org/articles/10.3389/fonc.2021.738832

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*, *2*(5), 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095

Chakraborty, S., Ahler, E., Simon, J. J., Fang, L., Potter, Z. E., Sitko, K. A., Stephany, J. J., Guttman, M., Fowler, D. M., & Maly, D. J. (2023). Profiling of drug resistance in Src kinase at scale uncovers a regulatory network coupling autoinhibition and catalytic domain dynamics. *Cell Chemical Biology*. https://doi.org/10.1016/j.chembiol.2023.08.005

Chakravarty, D., Gao, J., Phillips, S., Kundra, R., Zhang, H., Wang, J., Rudolph, J. E., Yaeger, R., Soumerai, T., Nissan, M. H., Chang, M. T., Chandarlapaty, S., Traina, T. A., Paik, P. K., Ho, A. L., Hantash, F. M., Grupe, A., Baxi, S. S., Callahan, M. K., … Schultz, N. (2017). OncoKB: A Precision Oncology Knowledge Base. JCO Precision Oncology, 1, 1–16. https://doi.org/10.1200/PO.17.00011
Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Cohen, P., Cross, D., & Jänne, P. A. (2021). Kinase drug discovery 20 years after imatinib: Progress and future directions. *Nature Reviews Drug Discovery*, *20*(7), Article 7. https://doi.org/10.1038/s41573-021-00195-4

Cooper, C. S., Park, M., Blair, D. G., Tainsky, M. A., Huebner, K., Croce, C. M., & Vande Woude, G. F. (1984). Molecular cloning of a new transforming gene from a chemically transformed human cell line. Nature, 311(5981), 29–33. https://doi.org/10.1038/311029a0

Cui, J. J. (2014). Targeting Receptor Tyrosine Kinase MET in Cancer: Small Molecule Inhibitors and Clinical Progress. *Journal of Medicinal Chemistry*, *57*(11), 4427–4453. https://doi.org/10.1021/jm401427c

Cui, J. J., Tran-Dubé, M., Shen, H., Nambu, M., Kung, P.-P., Pairish, M., Jia, L., Meng, J., Funk, L., Botrous, I., McTigue, M., Grodsky, N., Ryan, K., Padrique, E., Alton, G., Timofeevski, S., Yamazaki, S., Li, Q., Zou, H., … Edwards, M. P. (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). *Journal of Medicinal Chemistry*, *54*(18), 6342–6363. https://doi.org/10.1021/jm2007613

Daley, G. Q., & Baltimore, D. (1988). Transformation of an interleukin 3-dependent hematopoietic cell line by the chronic myelogenous leukemia-specific P210bcr/abl protein. Proceedings of the National Academy of Sciences, 85(23), 9312–9316. https://doi.org/10.1073/pnas.85.23.9312

Dieckhaus, H., Brocidiacono, M., Randolph, N., & Kuhlman, B. (2023). *Transfer learning to leverage larger datasets for improved prediction of protein stability changes* (p. 2023.07.27.550881). bioRxiv. https://doi.org/10.1101/2023.07.27.550881

Diedrich, K., Krause, B., Berg, O., & Rarey, M. (2023). PoseEdit: Enhanced ligand binding mode communication by interactive 2D diagrams. *Journal of Computer-Aided Molecular Design*, *37*(10), 491–503. https://doi.org/10.1007/s10822-023-00522-4

Dong, R., Peng, Z., Zhang, Y., & Yang, J. (2018). mTM-align: An algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics*, *34*(10), 1719–1725. https://doi.org/10.1093/bioinformatics/btx828

Dorighi, K. M., Zhu, A., Fortin, J.-P., Lo, J. H.-H., Sudhamsu, J., Wendorff, T. J., Durinck, S., Callow, M., Foster, S. A., & Haley, B. (2024). Accelerated drug-resistant variant discovery with an enhanced, scalable mutagenic base editor platform. *Cell Reports*, *43*(6). https://doi.org/10.1016/j.celrep.2024.114313

Duplaquet, L., Kherrouche, Z., Baldacci, S., Jamme, P., Cortot, A. B., Copin, M.-C., & Tulasne, D. (2018). The multiple paths towards MET receptor addiction in cancer. *Oncogene*, *37*(24), 3200–3215. https://doi.org/10.1038/s41388-018-0185-4

Eastman, P., Friedrichs, M. S., Chodera, J. D., Radmer, R. J., Bruns, C. M., Ku, J. P., Beauchamp, K. A., Lane, T. J., Wang, L.-P., Shukla, D., Tye, T., Houston, M., Stich, T., Klein, C., Shirts, M. R., & Pande, V. S. (2013). OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *Journal of Chemical Theory and Computation*, *9*(1), 461–469. https://doi.org/10.1021/ct300857j

Eathiraj, S., Palma, R., Volckova, E., Hirschi, M., France, D. S., Ashwell, M. A., & Chan, T. C. K. (2011). Discovery of a Novel Mode of Protein Kinase Inhibition Characterized by the Mechanism of Inhibition of Human Mesenchymal-epithelial Transition Factor (c-Met) Protein Autophosphorylation by ARQ 197. *Journal of Biological Chemistry*, *286*(23), 20666–20676. https://doi.org/10.1074/jbc.M110.213801

Eberhardt, J., Santos-Martins, D., Tillack, A. F., & Forli, S. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling*, *61*(8), 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203

Elizabeth A. Tovar, & Graveel, C. R. (2017). MET in human cancer: Germline and somatic mutations. *Annals of Translational Medicine*, *5*(10), 205–205. https://doi.org/10.21037/atm.2017.03.64

Estevam, G. O., Linossi, E. M., Macdonald, C. B., Espinoza, C. A., Michaud, J. M., Coyote-Maestas, W., Collisson, E. A., Jura, N., & Fraser, J. S. (2024). *Conserved regulatory motifs in the juxtamembrane domain* *and kinase N-lobe revealed through deep mutational scanning of the MET receptor tyrosine kinase domain* (p. 2023.08.03.551866). bioRxiv. https://doi.org/10.1101/2023.08.03.551866

Fernandes, M., Jamme, P., Cortot, A. B., Kherrouche, Z., & Tulasne, D. (2021). When the MET receptor kicks in to resist targeted therapies. *Oncogene*, *40*(24), 4061–4078. https://doi.org/10.1038/s41388-021-01835-0

Frampton, G. M., Ali, S. M., Rosenzweig, M., Chmielecki, J., Lu, X., Bauer, T. M., Akimov, M., Bufill, J. A., Lee, C., Jentz, D., Hoover, R., Ou, S.-H. I., Salgia, R., Brennan, T., Chalmers, Z. R., Jaeger, S., Huang, A., Elvin, J. A., Erlich, R., … Miller, V. A. (2015). Activation of MET via Diverse Exon 14 Splicing Alterations Occurs in Multiple Tumor Types and Confers Clinical Sensitivity to MET Inhibitors. *Cancer Discovery*, *5*(8), 850–859. https://doi.org/10.1158/2159-8290.CD-15-0285

Fujino, T., Kobayashi, Y., Suda, K., Koga, T., Nishino, M., Ohara, S., Chiba, M., Shimoji, M., Tomizawa, K., Takemoto, T., & Mitsudomi, T. (2019). Sensitivity and Resistance of MET Exon 14 Mutations in Lung Cancer to Eight MET Tyrosine Kinase Inhibitors In Vitro. *Journal of Thoracic Oncology*, *14*(10), 1753–1765. https://doi.org/10.1016/j.jtho.2019.06.023

Fujino, T., Suda, K., Koga, T., Hamada, A., Ohara, S., Chiba, M., Shimoji, M., Takemoto, T., Soh, J., & Mitsudomi, T. (2022). Foretinib can overcome common on-target resistance mutations after capmatinib/tepotinib treatment in NSCLCs with MET exon 14 skipping mutation. *Journal of Hematology & Oncology*, *15*(1), 79. https://doi.org/10.1186/s13045-022-01299-z

Fujino, T., Suda, K., & Mitsudomi, T. (2021). Lung Cancer with MET exon 14 Skipping Mutation: Genetic Feature, Current Treatments, and Future Challenges. *Lung Cancer: Targets and Therapy*, *12*, 35–50. https://doi.org/10.2147/LCTT.S269307

Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., & Caves, L. S. D. (2006). Bio3d: An R package for the comparative analysis of protein structures. Bioinformatics, 22(21), 2695–2696. https://doi.org/10.1093/bioinformatics/btl461

Guérin, C. M., Vinchent, A., Fernandes, M., Damour, I., Laratte, A., Tellier, R., Estevam, G. O., Meneboo, J.-P., Villenet, C., Descarpentries, C., Fraser, J. S., Figeac, M., Cortot, A. B., Rouleau, E., & Tulasne, D. (2023). *MET variants with activating N-lobe mutations identified in hereditary papillary renal cell carcinomas still require ligand stimulation* (p. 2023.11.03.565283). bioRxiv. https://doi.org/10.1101/2023.11.03.565283

Jones, E. M., Lubock, N. B., Venkatakrishnan, A., Wang, J., Tseng, A. M., Paggi, J. M., Latorraca, N. R., Cancilla, D., Satyadi, M., Davis, J. E., Babu, M. M., Dror, R. O., & Kosuri, S. (2020). Structural and functional characterization of G protein–coupled receptors with deep mutational scanning. *eLife*, *9*, e54895. https://doi.org/10.7554/eLife.54895

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), Article 7873. https://doi.org/10.1038/s41586-021-03819-2

Katayama, R., Aoyama, A., Yamori, T., Qi, J., Oh-hara, T., Song, Y., Engelman, J. A., & Fujita, N. (2013). Cytotoxic Activity of Tivantinib (ARQ 197) Is Not Due Solely to c-MET Inhibition. *Cancer Research*, *73*(10), 3087–3096. https://doi.org/10.1158/0008-5472.CAN-12-3256

Koga, T., Suda, K., & Mitsudomi, T. (2022). Utility of the Ba/F3 cell system for exploring on-target mechanisms of resistance to targeted therapies for lung cancer. Cancer Science, 113(3), 815–827. https://doi.org/10.1111/cas.15263

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, *10*(1), 168. https://doi.org/10.1186/1471-2105-10-168

Li, A., Yang, J., Zhang, X., Zhang, Z., Su, J., Gou, L., Bai, Y., Zhou, Q., Yang, Z., Han-Zhang, H., Zhong, W.-Z., Chuai, S., Zhang, Q., Xie, Z., Gao, H., Chen, H., Wang, Z., Wang, Z., Yang, X., … Wu, Y. (2017). Acquired MET Y1248H and D1246N Mutations Mediate Resistance to MET Inhibitors in Non–Small Cell Lung Cancer. *Clinical Cancer Research*, *23*(16), 4929–4937. https://doi.org/10.1158/1078-0432.CCR-16-3273

Lin, A., Giuliano, C. J., Palladino, A., John, K. M., Abramowicz, C., Yuan, M. L., Sausville, E. L., Lukow, D. A., Liu, L., Chait, A. R., Galluzzo, Z. C., Tucker, C., & Sheltzer, J. M. (2019). Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Science Translational Medicine*, *11*(509), eaaw8412. https://doi.org/10.1126/scitranslmed.aaw8412

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., & Rives, A. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, *379*(6637), 1123–1130. https://doi.org/10.1126/science.ade2574

Lu, X., Peled, N., Greer, J., Wu, W., Choi, P., Berger, A. H., Wong, S., Jen, K.-Y., Seo, Y., Hann, B., Brooks, A., Meyerson, M., & Collisson, E. A. (2017). MET Exon 14 Mutation Encodes an Actionable Therapeutic Target in Lung Adenocarcinoma. *Cancer Research*, *77*(16), 4498–4505. https://doi.org/10.1158/0008-5472.CAN-16-1944

Ma, P. C., Kijima, T., Maulik, G., Fox, E. A., Sattler, M., Griffin, J. D., Johnson, B. E., & Salgia, R. (2003). c-MET Mutational Analysis in Small Cell Lung Cancer: Novel Juxtamembrane Domain Mutations Regulating Cytoskeletal Functions1. Cancer Research, 63(19), 6272–6281.

Mak, H. H. L., Peschard, P., Lin, T., Naujokas, M. A., Zuo, D., & Park, M. (2007). Oncogenic activation of the Met receptor tyrosine kinase fusion protein, Tpr–Met, involves exclusion from the endocytic degradative pathway. Oncogene, 26(51), 7213–7221. https://doi.org/10.1038/sj.onc.1210522

Meier, J., Rao, R., Verkuil, R., Liu, J., Sercu, T., & Rives, A. (2021). *Language models enable zero-shot prediction of the effects of mutations on protein function* (p. 2021.07.09.450648). bioRxiv. https://doi.org/10.1101/2021.07.09.450648

Michaelides, I. N., Collie, G. W., Börjesson, U., Vasalou, C., Alkhatib, O., Barlind, L., Cheung, T., Dale, I. L., Embrey, K. J., Hennessy, E. J., Khurana, P., Koh, C. M., Lamb, M. L., Liu, J., Moss, T. A., O'Neill, D. J., Phillips, C., Shaw, J., Snijder, A., … Yang, W. (2023). Discovery and Optimization of the First ATP Competitive Type-III c-MET Inhibitor. *Journal of Medicinal Chemistry*, *66*(13), 8782–8807. https://doi.org/10.1021/acs.jmedchem.3c00401

Michieli, P., & Di Nicolantonio, F. (2013). Tivantinib—A cytotoxic drug in MET inhibitor's clothes? *Nature Reviews Clinical Oncology*, *10*(7), Article 7. https://doi.org/10.1038/nrclinonc.2013.86

Mingione, V. R., Paung, Y., Outhwaite, I. R., & Seeliger, M. A. (2023). Allosteric regulation and inhibition of protein kinases. *Biochemical Society Transactions*, *51*(1), 373–385. https://doi.org/10.1042/BST20220940

Munshi, N., Jeay, S., Li, Y., Chen, C.-R., France, D. S., Ashwell, M. A., Hill, J., Moussa, M. M., Leggett, D. S., & Li, C. J. (2010). ARQ 197, a Novel and Selective Inhibitor of the Human c-Met Receptor Tyrosine Kinase with Antitumor Activity. *Molecular Cancer Therapeutics*, *9*(6), 1544–1553. https://doi.org/10.1158/1535-7163.MCT-09-1173

Pal, K., Bandyopadhyay, A., Zhou, X. E., Xu, Q., Marciano, D. P., Brunzelle, J. S., Yerrum, S., Griffin, P. R., Vande Woude, G., Melcher, K., & Xu, H. E. (2017). Structural Basis of TPR-Mediated Oligomerization and Activation of Oncogenic Fusion Kinases. Structure, 25(6), 867-877.e3. https://doi.org/10.1016/j.str.2017.04.015

Park, M., Dean, M., Cooper, C. S., Schmidt, M., O'Brien, S. J., Blair, D. G., & Vande Woude, G. F. (1986). Mechanism of met oncogene activation. Cell, 45(6), 895–904. https://doi.org/10.1016/0092-8674(86)90564-7

Pecci, F., Nakazawa, S., Ricciuti, B., Harada, G., Lee, J. K., Alessi, J. V., Barrichello, A., Vaz, V. R., Lamberti, G., Di Federico, A., Gandhi, M. M., Gazgalis, D., Feng, W. W., Jiang, J., Baldacci, S., Locquet, M.-A., Gottlieb, F. H., Chen, M. F., Lee, E., … Awad, M. M. (2024). Activating Point Mutations in the MET Kinase Domain Represent a Unique Molecular Subset of Lung Cancer and Other Malignancies Targetable with MET Inhibitors. Cancer Discovery, OF1–OF17. https://doi.org/10.1158/2159-8290.CD-23-1217

Persky, N. S., Hernandez, D., Do Carmo, M., Brenan, L., Cohen, O., Kitajima, S., Nayar, U., Walker, A., Pantel, S., Lee, Y., Cordova, J., Sathappa, M., Zhu, C., Hayes, T. K., Ram, P., Pancholi, P., Mikkelsen, T. S., Barbie, D. A., Yang, X., … Johannessen, C. M. (2020). Defining the landscape of ATP-competitive inhibitor resistance residues in protein kinases. Nature Structural & Molecular Biology, 27(1), 92–104. https://doi.org/10.1038/s41594-019-0358-z

Peschard, P., Fournier, T. M., Lamorte, L., Naujokas, M. A., Band, H., Langdon, W. Y., & Park, M. (2001). Mutation of the c-Cbl TKB Domain Binding Site on the Met Receptor Tyrosine Kinase Converts It into a Transforming Protein. Molecular Cell, 8(5), 995–1004. https://doi.org/10.1016/S1097-2765(01)00378-1

Rao, J., Xin, R., Macdonald, C., Howard, M. K., Estevam, G. O., Yee, S. W., Wang, M., Fraser, J. S., Coyote-Maestas, W., & Pimentel, H. (2024). Rosace: A robust deep mutational scanning analysis framework employing position and mean-variance shrinkage. *Genome Biology*, *25*(1), 138. https://doi.org/10.1186/s13059-024-03279-7

Recondo, G., Bahcall, M., Spurr, L. F., Che, J., Ricciuti, B., Leonardi, G. C., Lo, Y.-C., Li, Y. Y., Lamberti, G., Nguyen, T., Milan, M. S. D., Venkatraman, D., Umeton, R., Paweletz, C. P., Albayrak, A., Cherniack, A. D., Price, K. S., Fairclough, S. R., Nishino, M., … Awad, M. M. (2020). Molecular Mechanisms of Acquired Resistance to MET Tyrosine Kinase Inhibitors in Patients with MET Exon 14–Mutant NSCLC. *Clinical Cancer Research*, *26*(11), 2615–2625. https://doi.org/10.1158/1078-0432.CCR-19-3608

Recondo, G., Che, J., Jänne, P. A., & Awad, M. M. (2020). Targeting *MET* Dysregulation in Cancer. *Cancer Discovery*, *10*(7), 922–934. https://doi.org/10.1158/2159-8290.CD-19-1446

Recondo, G., Facchinetti, F., Olaussen, K. A., Besse, B., & Friboulet, L. (2018). Making the first move in EGFR-driven or ALK-driven NSCLC: First-generation or next-generation TKI? *Nature Reviews Clinical Oncology*, *15*(11), 694–708. https://doi.org/10.1038/s41571-018-0081-4

Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., & Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, *118*(15), e2016239118. https://doi.org/10.1073/pnas.2016239118

Rodrigues, G. A., & Park, M. (1993). Dimerization mediated through a leucine zipper activates the oncogenic potential of the met receptor tyrosine kinase. Molecular and Cellular Biology, 13(11), 6711–6722.

Rubin, A. F., Gelman, H., Lucas, N., Bajjalieh, S. M., Papenfuss, A. T., Speed, T. P., & Fowler, D. M. (2017). A statistical framework for analyzing deep mutational scanning data. *Genome Biology*, *18*(1), 150. https://doi.org/10.1186/s13059-017-1272-5

Sadybekov, A. V., & Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature*, *616*(7958), Article 7958. https://doi.org/10.1038/s41586-023-05905-z

Santarpia, M., Massafra, M., Gebbia, V., D'Aquino, A., Garipoli, C., Altavilla, G., & Rosell, R. (2021). A narrative review of MET inhibitors in non-small cell lung cancer with MET exon 14 skipping mutations. *Translational Lung Cancer Research*, *10*(3). https://doi.org/10.21037/tlcr-20-1113

Smyth, E. C., Sclafani, F., & Cunningham, D. (2014). Emerging molecular targets in oncology: Clinical potential of MET/hepatocyte growth-factor inhibitors. *OncoTargets and Therapy*, *7*, 1001–1014. https://doi.org/10.2147/OTT.S44941

Suehnholz, S. P., Nissan, M. H., Zhang, H., Kundra, R., Nandakumar, S., Lu, C., Carrero, S., Dhaneshwar, A., Fernandez, N., Xu, B. W., Arcila, M. E., Zehir, A., Syed, A., Brannon, A. R., Rudolph, J. E., Paraiso, E., Sabbatini, P. J., Levine, R. L., Dogan, A., … Chakravarty, D. (2024). Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. Cancer Discovery, 14(1), 49–65. https://doi.org/10.1158/2159-8290.CD-23-0467

Tanizaki, J., Okamoto, I., Okamoto, K., Takezawa, K., Kuwata, K., Yamaguchi, H., & Nakagawa, K. (2011). MET Tyrosine Kinase Inhibitor Crizotinib (PF-02341066) Shows Differential Antitumor Effects in Non-small Cell Lung Cancer According to MET Alterations. *Journal of Thoracic Oncology*, *6*(10), 1624–1631. https://doi.org/10.1097/JTO.0b013e31822591e9

Tiedt, R., Degenkolbe, E., Furet, P., Appleton, B. A., Wagner, S., Schoepfer, J., Buck, E., Ruddy, D. A., Monahan, J. E., Jones, M. D., Blank, J., Haasen, D., Drueckes, P., Wartmann, M., McCarthy, C., Sellers, W. R., & Hofmann, F. (2011). A Drug Resistance Screen Using a Selective MET Inhibitor Reveals a Spectrum of Mutations That Partially Overlap with Activating Mutations Found in Cancer Patients. *Cancer Research*, *71*(15), 5255–5264. https://doi.org/10.1158/0008-5472.CAN-10-4433

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334

Vigna, E., Gramaglia, D., Longati, P., Bardelli, A., & Comoglio, P. M. (1999). Loss of the exon encoding the juxtamembrane domain is essential for the oncogenic activation of TPR-MET. Oncogene, 18(29), 4275–4281. https://doi.org/10.1038/sj.onc.1202791

Wang, C., & Lu, X. (2023). Targeting MET: Discovery of Small Molecule Inhibitors as Non-Small Cell Lung Cancer Therapy. *Journal of Medicinal Chemistry*, *66*(12), 7670–7697. https://doi.org/10.1021/acs.jmedchem.3c00028

Warmuth, M., Kim, S., Gu, X., Xia, G., & Adrián, F. (2007). Ba/F3 cells and their use in kinase drug discovery. Current Opinion in Oncology, 19(1), 55–60. https://doi.org/10.1097/CCO.0b013e328011a25f

Wójcikowski, M., Ballester, P. J., & Siedlecki, P. (2017). Performance of machine-learning scoring functions in structure-based virtual screening. *Scientific Reports*, *7*(1), 46710. https://doi.org/10.1038/srep46710

Wood, G. E., Hockings, H., Hilton, D. M., & Kermorgant, S. (2021). The role of MET in chemotherapy resistance. *Oncogene*, *40*(11), 1927–1941. https://doi.org/10.1038/s41388-020-01577-5

Yao, Y., Yang, H., Zhu, B., Wang, S., Pang, J., Wu, X., Xu, Y., Zhang, J., Zhang, J., Ou, Q., Tian, H., & Zhao, Z. (2023). Mutations in the MET tyrosine kinase domain and resistance to tyrosine kinase inhibitors in non-small-cell lung cancer. *Respiratory Research*, *24*(1), 28. https://doi.org/10.1186/s12931-023-02329-1

Zhang, S., Krieger, J. M., Zhang, Y., Kaya, C., Kaynak, B., Mikulska-Ruminska, K., Doruker, P., Li, H., & Bahar, I. (2021). ProDy 2.0: Increased scale and scope after 10 years of protein dynamics modelling with Python. *Bioinformatics*, *37*(20), 3657–3659. https://doi.org/10.1093/bioinformatics/btab187

Zuccotto, F., Ardini, E., Casale, E., & Angiolini, M. (2010). Through the “Gatekeeper Door”: Exploiting the Active Kinase Conformation. *Journal of Medicinal Chemistry*, *53*(7), 2681–2694. https://doi.org/10.1021/jm901443h

Supplemental information

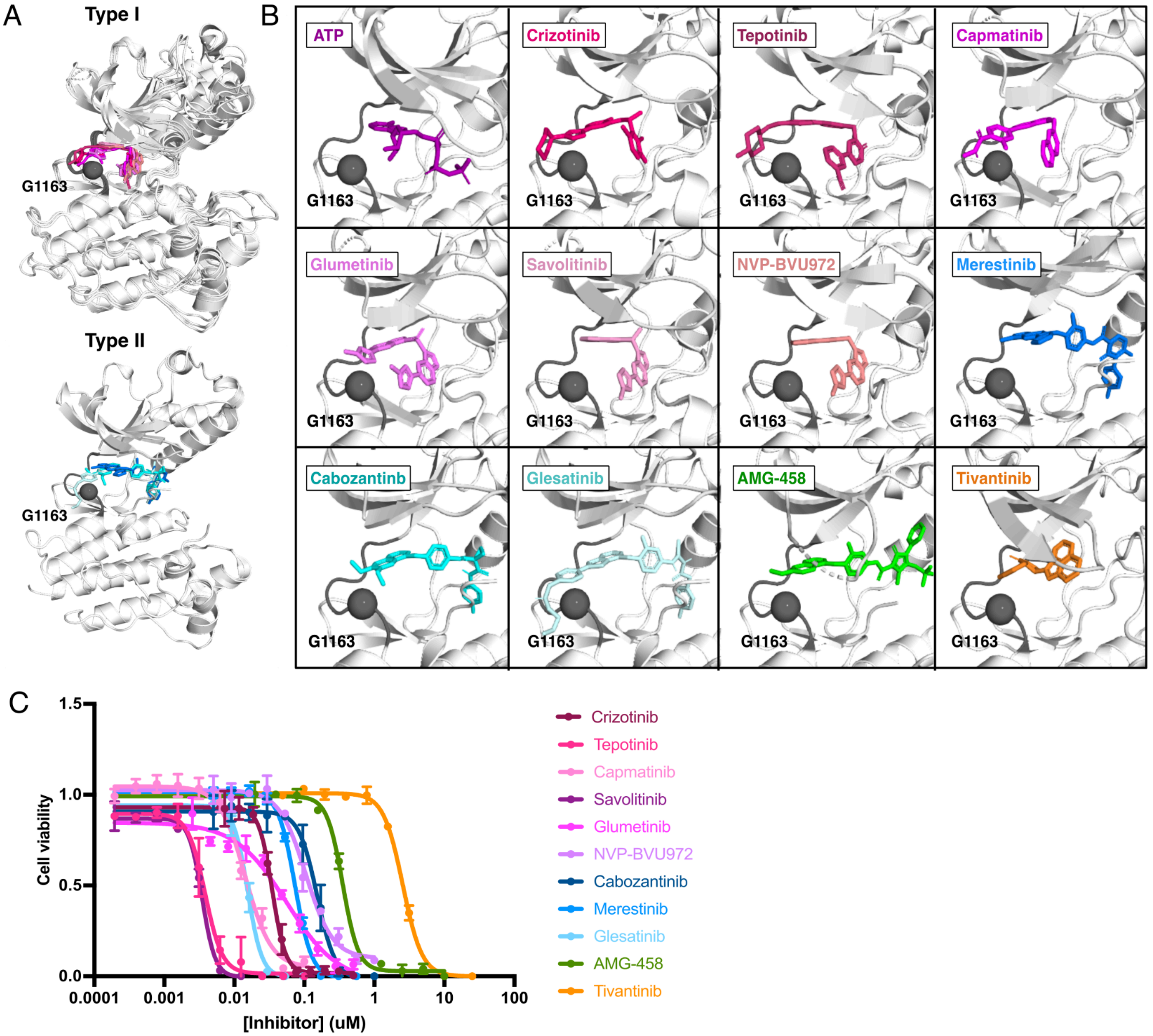

A
Type I
G1163
Type II
G1163
B
ATP
G1163
Crizotinib
G1163
Tepotinib
G1163
Capmatinib
G1163
Glumetinib
G1163
Savolitinib
G1163
NVP-BVU972
G1163
Merestinib
G1163
Cabozantinib
G1163
Glesatinib
G1163
AMG-458
G1163
Tivantinib
G1163
C
Cell viability
1.5
1.0
0.5
0.0
0.0001
0.001
0.01
0.1
1
10
100
[Inhibitor] (uM)
Crizotinib
Tepotinib
Capmatinib
Savolitinib
Glumetinib
NVP-BVU972
Cabozantinib
Merestinib
Glesatinib
AMG-458
Tivantinib

Figure 1 - figure supplement 1. Structural inhibitor classification and dose-response determination. (A) Type I (crizotinib, 2WGJ; tepotinib, 4R1V; capmatinib; savolitinib, 6SDE; NVP-BVU972, 3QTI) and type II (merestinib, 4EEV; cabozantinib; glesatinib) inhibitor-bound MET kinase domain structures globally aligned. Hinge (gray) and G1163 (represented as a sphere) are highlighted to show the kinase domain solvent-front relative to each inhibitor. Inhibitors lacking experimental structures (capmatinib, cabozantinib, glumetinib, and glesatinib) were docked onto a representative type I (PDB 2WGJ) and type II (4EEV) structure through AutoDock Vina **(Eberheart et al., 2021; Trott et al., 2010)**. (B) Solvent-front and G1163 highlighted relative to the ATP-bound kinase domain crystal structure (3DKC) and all inhibitors screened. (C) Dose-response curves for each inhibitor against the TPR-fusion MET and METΔEx14 intracellular domains, stably expressed in Ba/F3 cells.

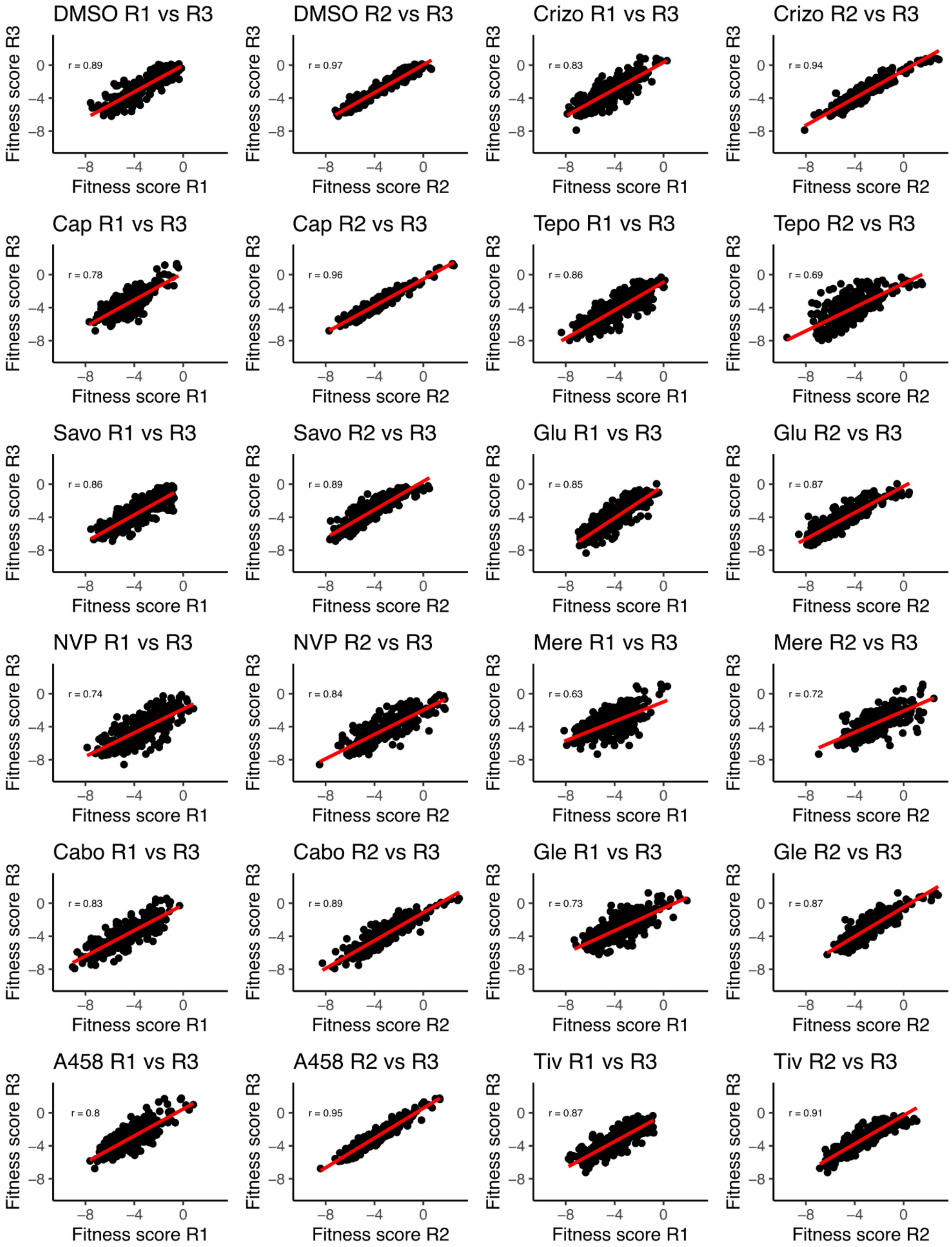


Figure 1 - figure supplement 2. Correlation analysis of the MET kinase domain site saturation mutagenesis library across replicates and conditions. Replicate correlation analysis for each inhibitor for both the TPR-fusion MET background scores with Enrich2.

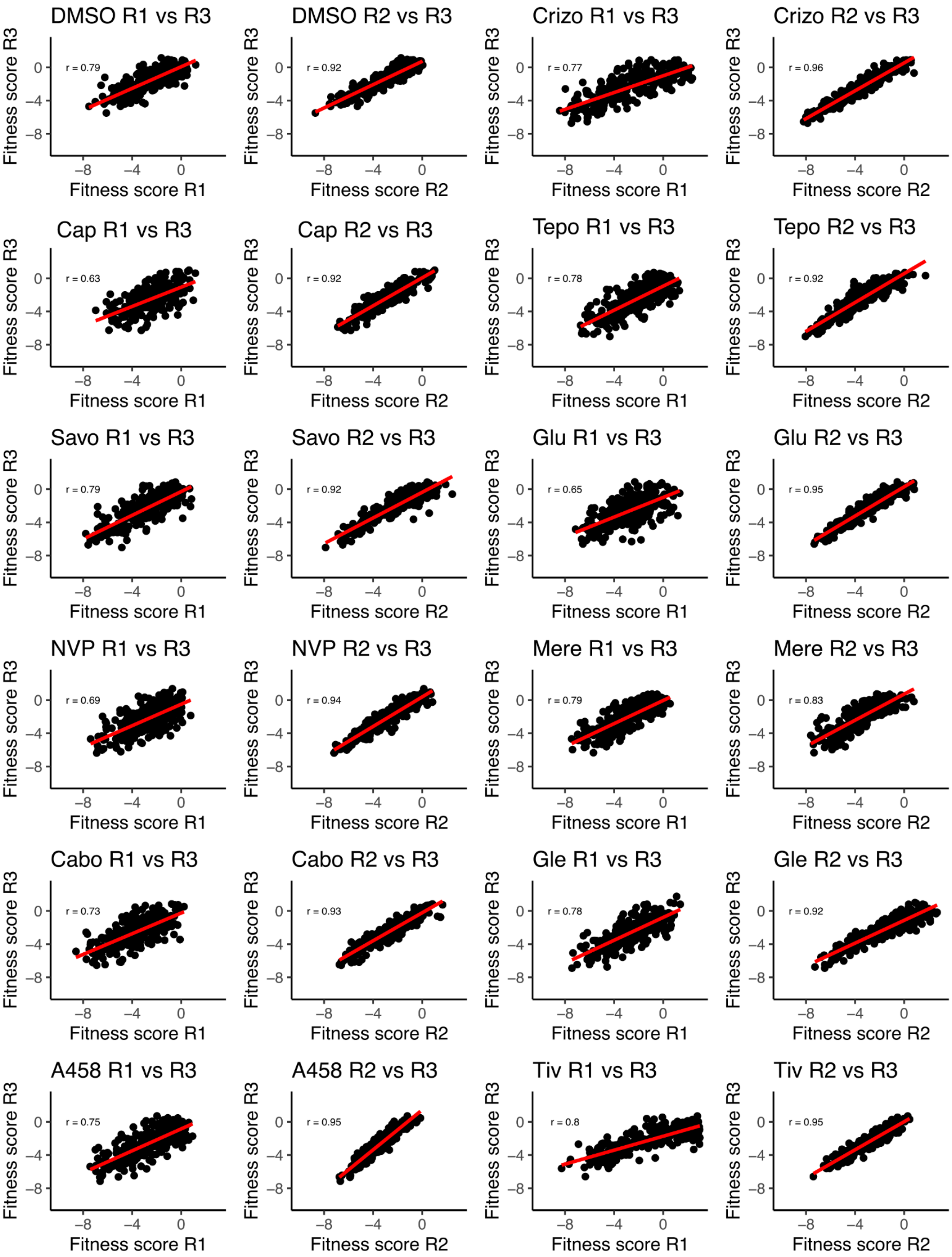
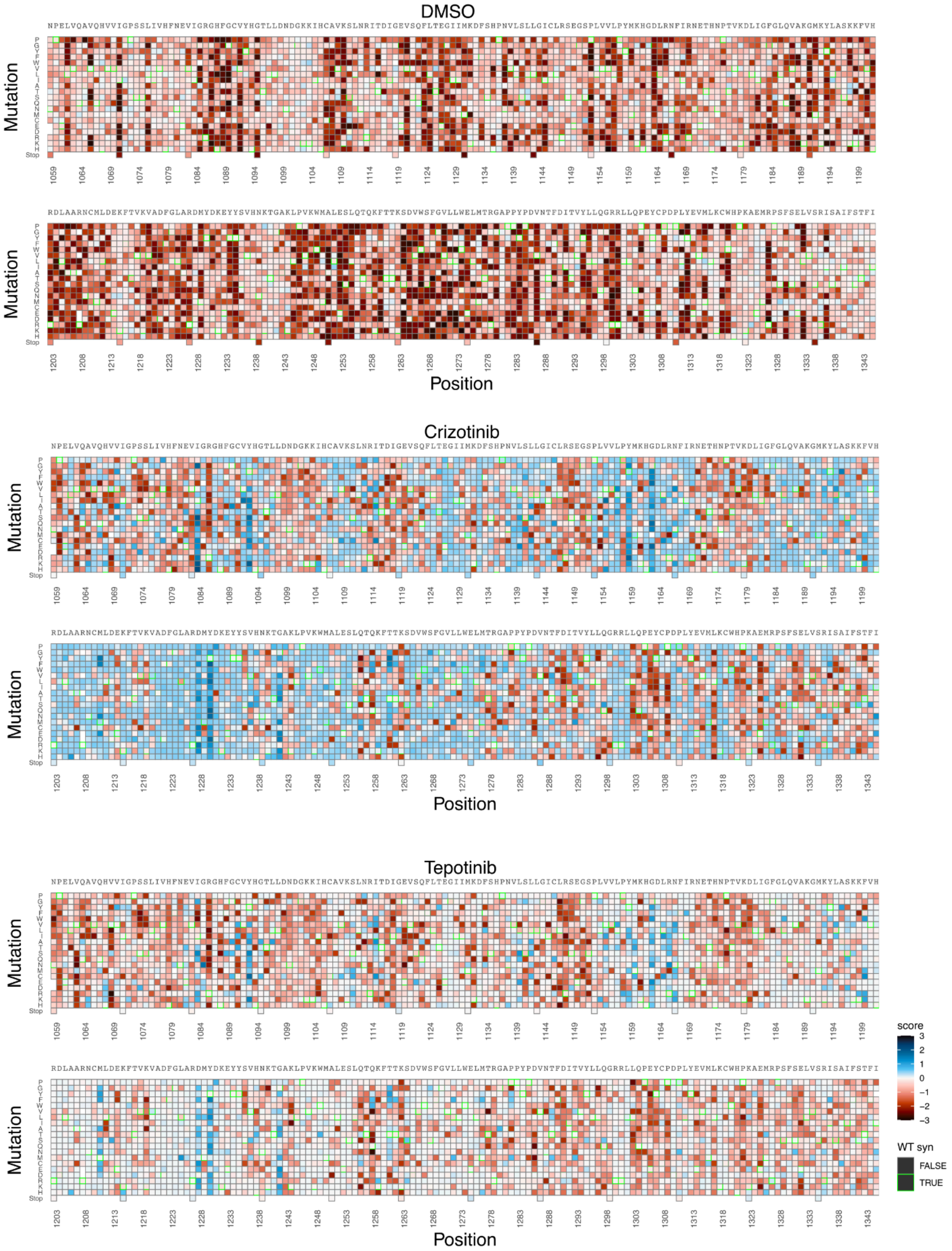
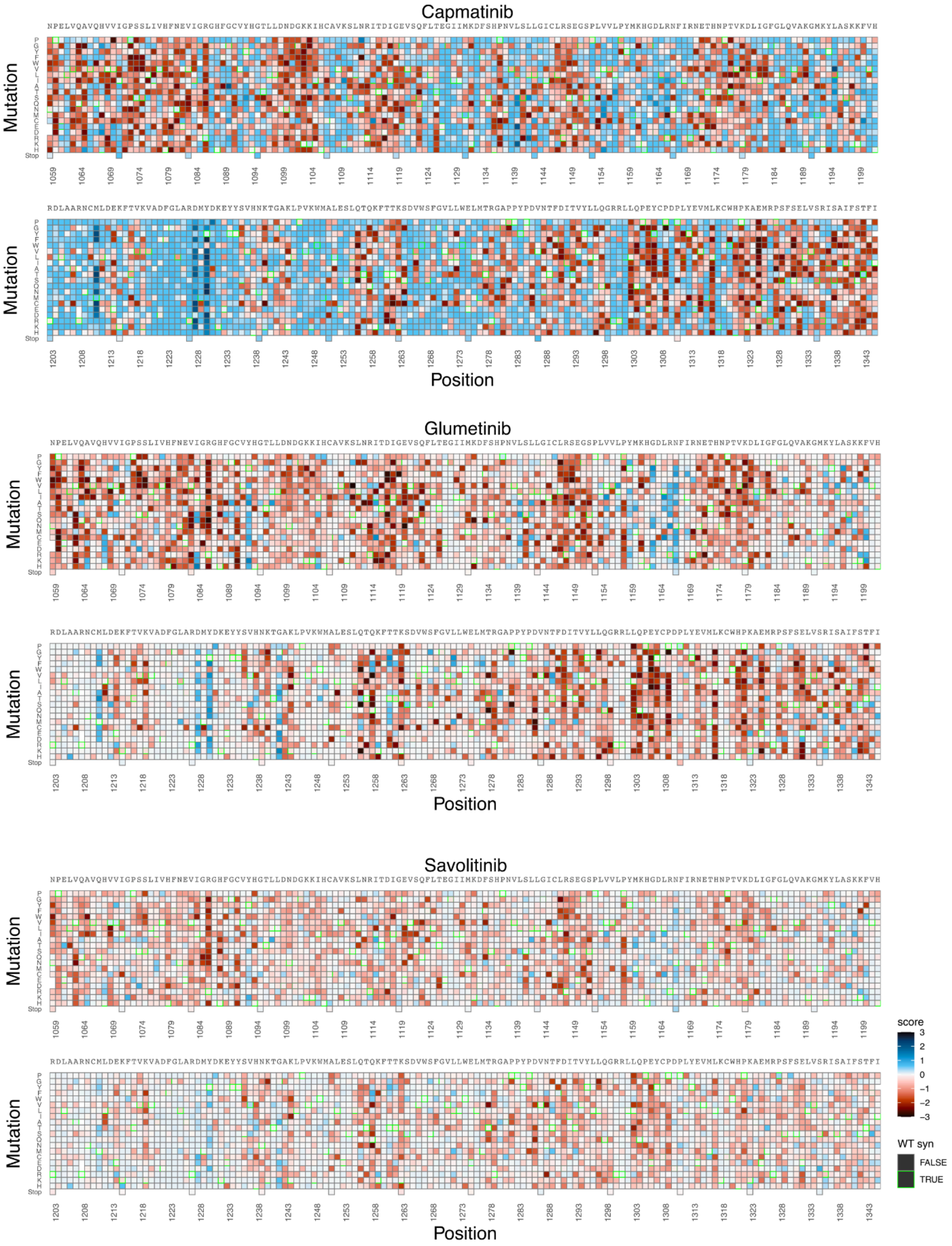


Figure 1 - figure supplement 3. Correlation analysis of the MET∆Ex14 kinase domain site saturation mutagenesis library across replicates and conditions. Replicate correlation analysis for each inhibitor for the TPR-fusion MET∆Ex14 background score with Enrich2.

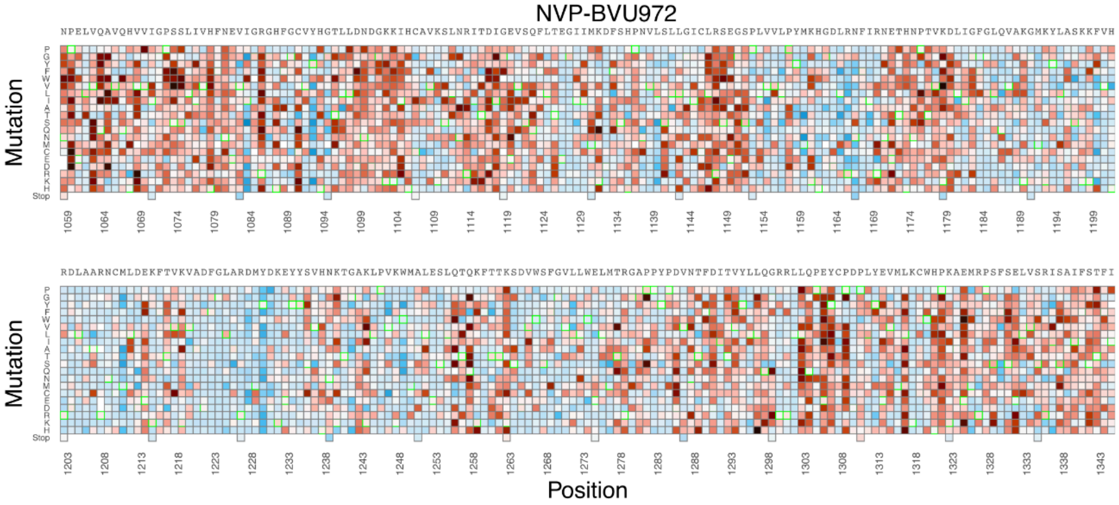
MET kinase domain heatmaps (1/4)

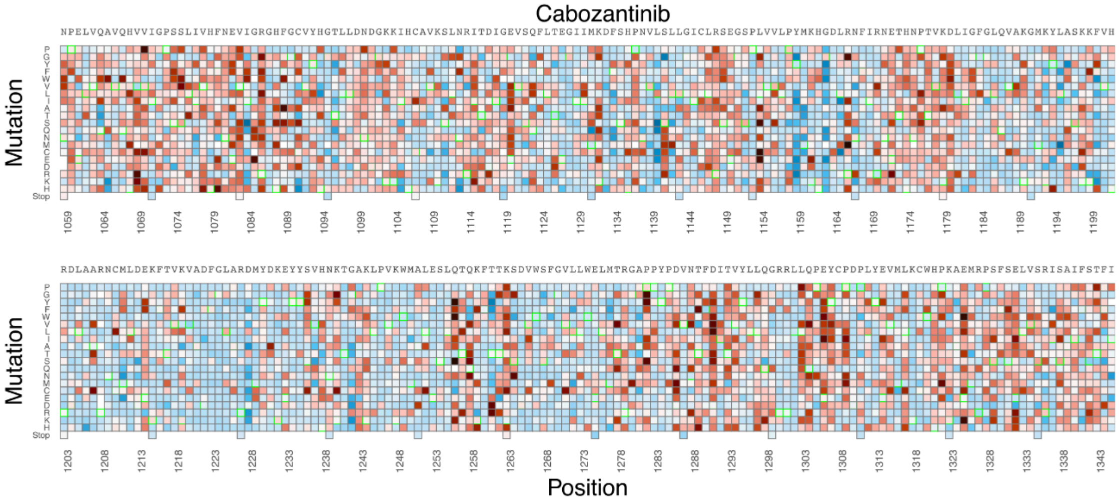


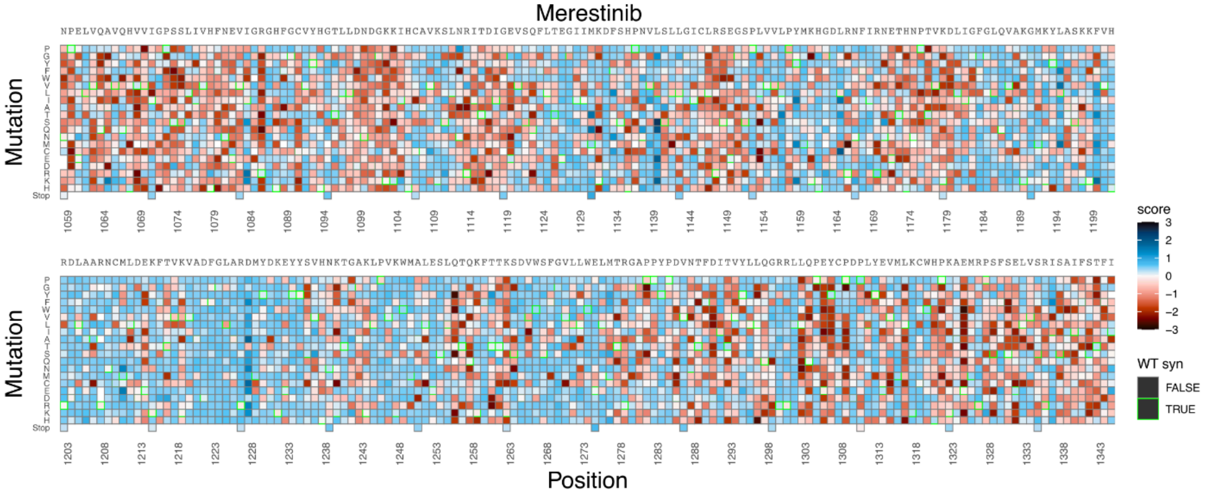
MET kinase domain heatmaps (2/4)



MET kinase domain heatmaps (3/4)

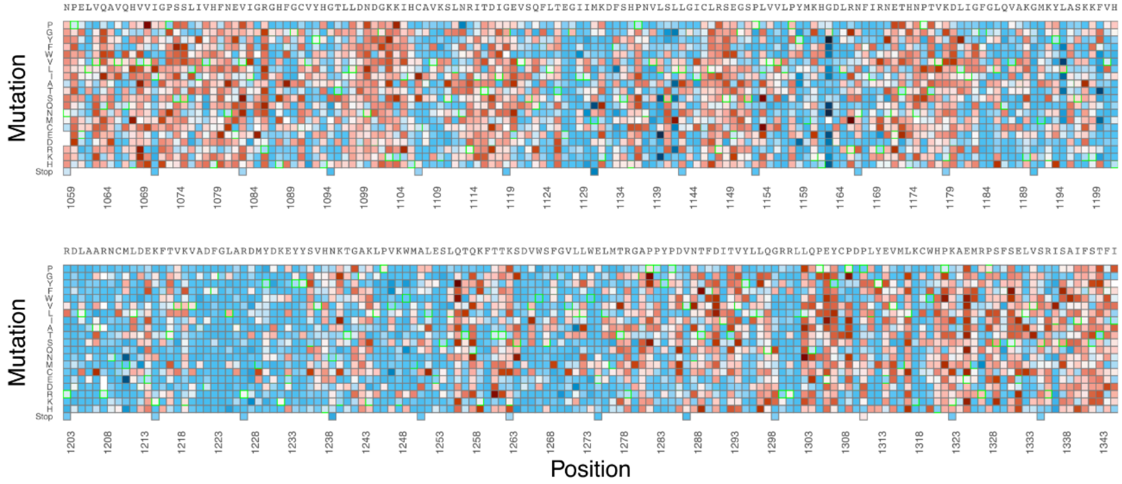

NVP-BVU972


Cabozantinib

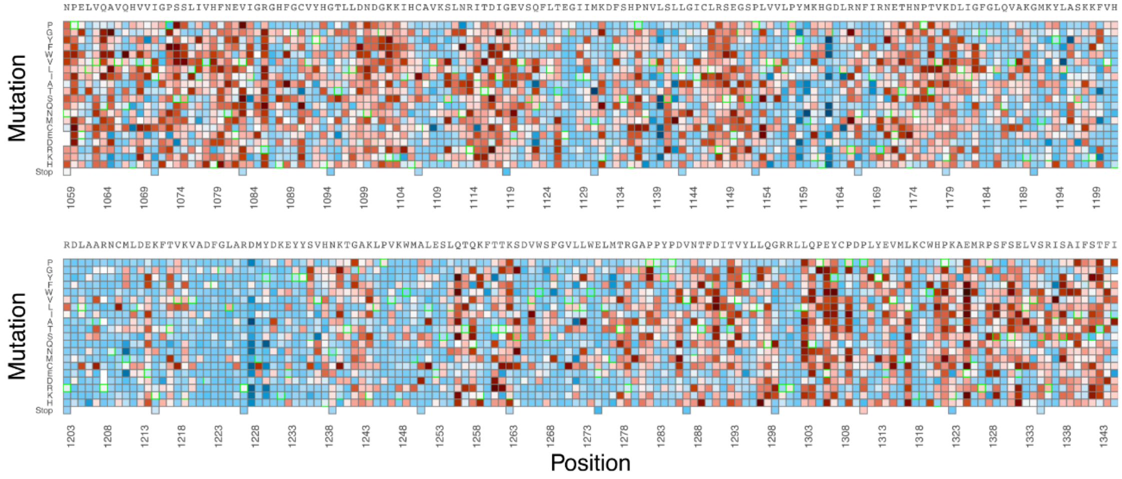

Merestinib

MET kinase domain heatmaps (4/4)

Glesatinib



AMG-458



Tivantinib

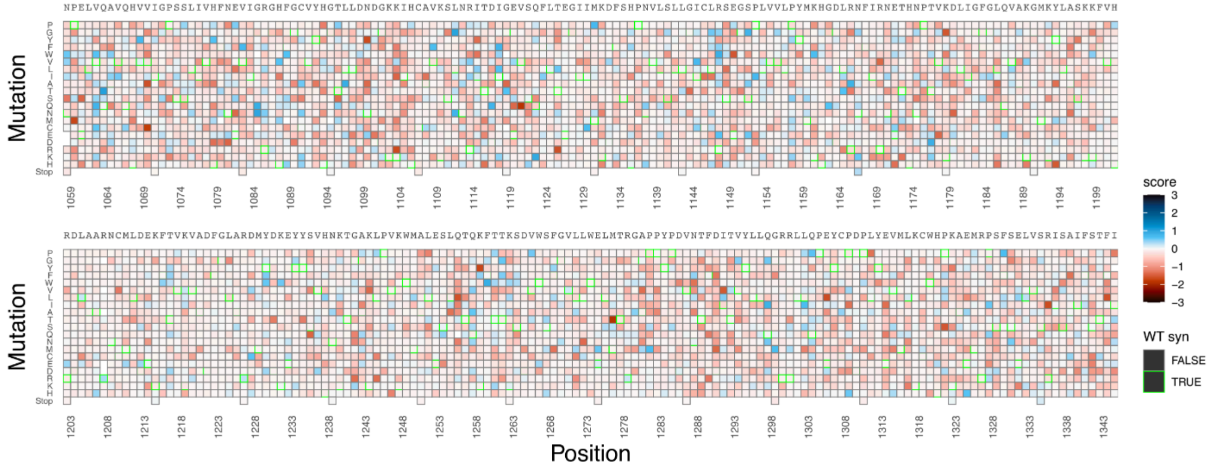
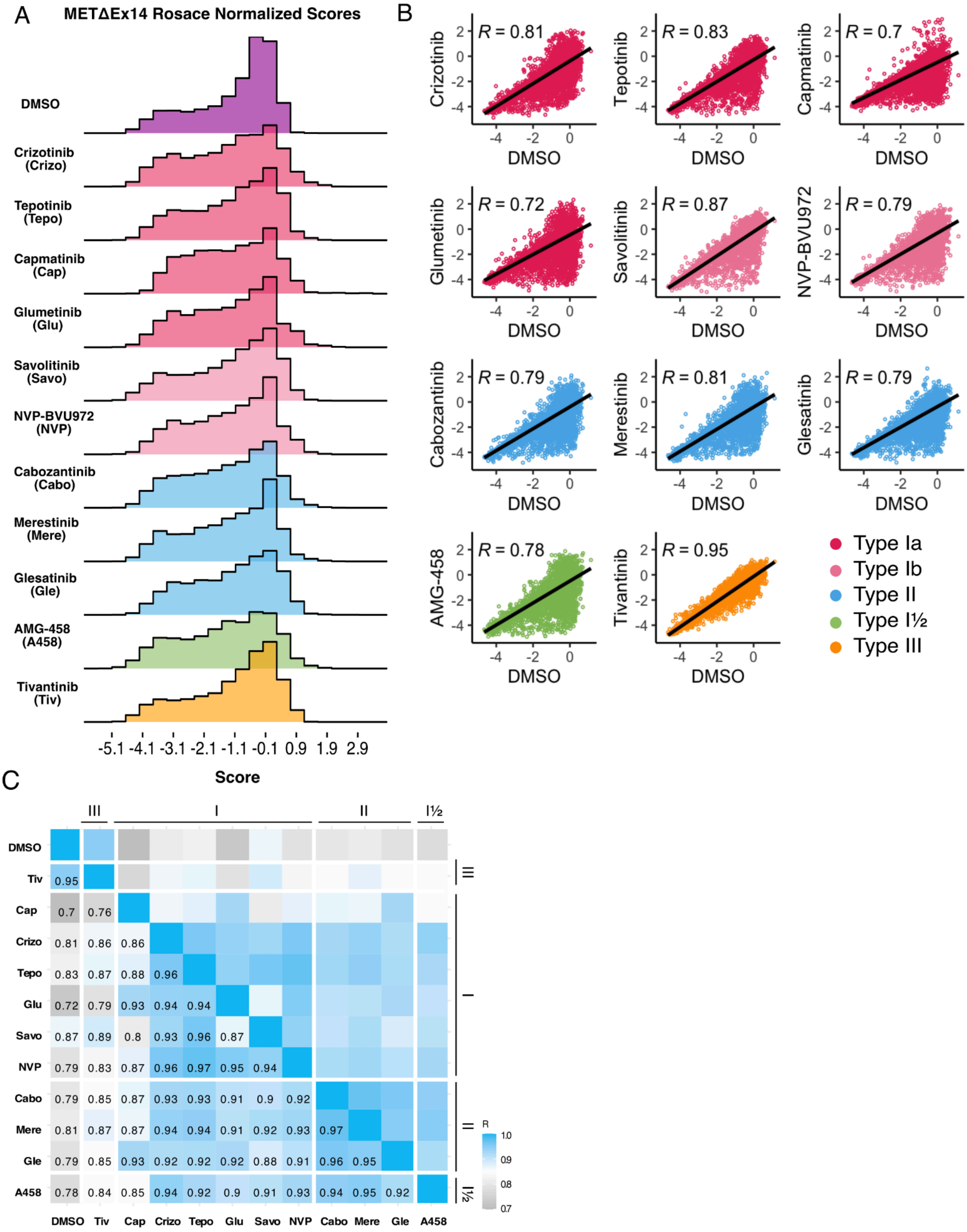
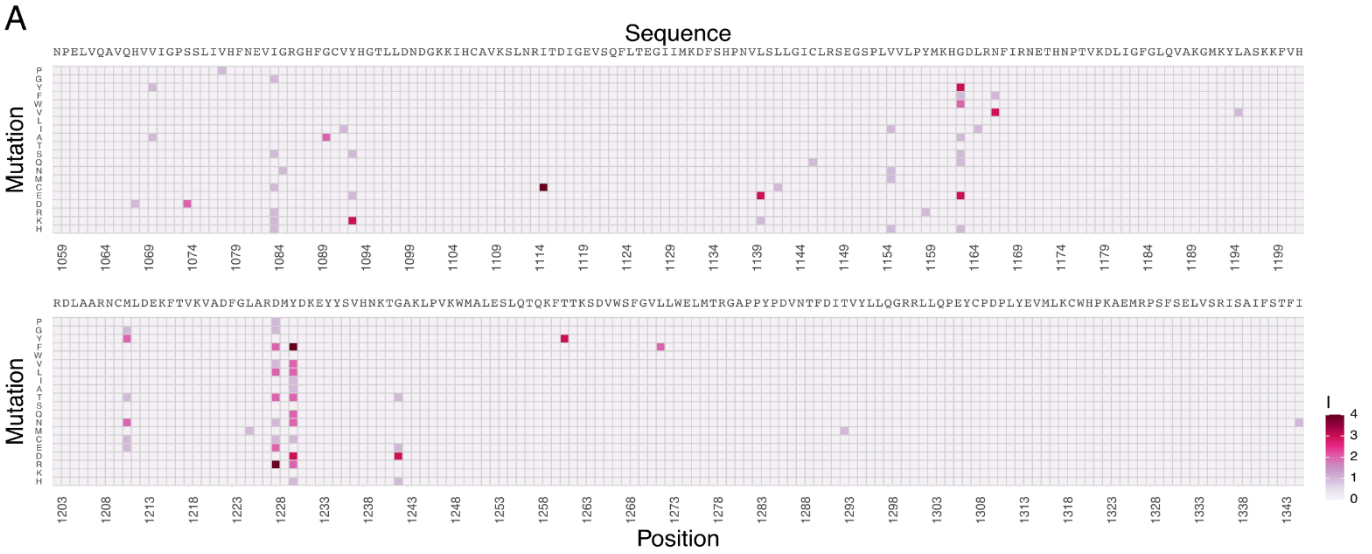


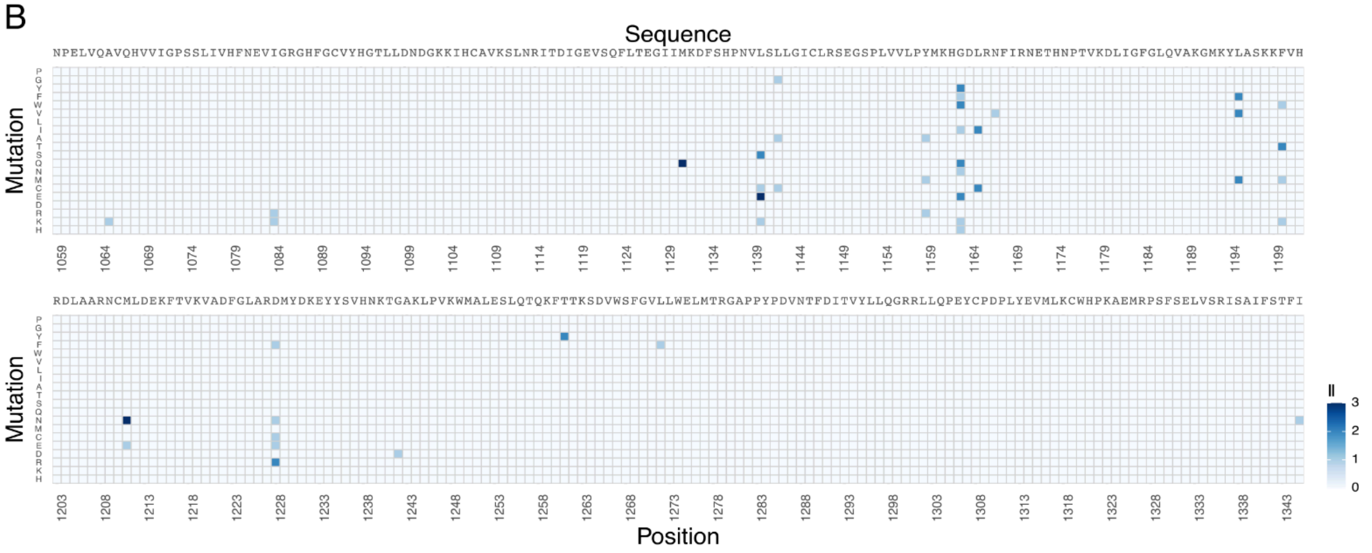
Figure 1 - figure supplement 4 . Fitness landscapes of the MET kinase domain against a panel of 11 inhibitors. Heatmap for the DMSO control condition and all inhibitor fitness scores from Rosace, subtracted from DMSO for >99% of MET kinase domain variants in the full intracellular domain background in the context of the TPR-fusion.



	DMSO	Tiv	Cap	Crizo	Tepo	Glu	Savo	NVP	Cabo	Mere	Gle	A458
DMSO												
Tiv	0.95											
Cap	0.7	0.76										
Crizo	0.81	0.86	0.86									
Tepo	0.83	0.87	0.88	0.96								
Glu	0.72	0.79	0.93	0.94	0.94							
Savo	0.87	0.89	0.8	0.93	0.96	0.87						
NVP	0.79	0.83	0.87	0.96	0.97	0.95	0.94					
Cabo	0.79	0.85	0.87	0.93	0.93	0.91	0.9	0.92				
Mere	0.81	0.87	0.87	0.94	0.94	0.91	0.92	0.93	0.97			
Gle	0.79	0.85	0.93	0.92	0.92	0.92	0.88	0.91	0.96	0.95		
A458	0.78	0.84	0.85	0.94	0.92	0.9	0.91	0.93	0.94	0.95	0.92	

Figure 2 - figure supplement 5. Mutational landscape of the METΔEx14 kinase domain under 11 ATP-competitive inhibitor selection. (A) Distributions of all variants (**wild-type** synonymous, early stop, and missense) for each condition, scored with Rosace and normalized to the growth rate of the DMSO control population. (B) Correlation plots for all mutational fitness scores for each drug against DMSO, fitted with a linear regression and Pearson's R value displayed. (C) Heatmap showing the Pearson's R correlation for each condition against each other, annotated by condition and inhibitor type. Correlations are colored according to a scale bar from gray to blue (low to high correlation).


A
Sequence
NPELVQAVQHVVIGPSSLIVHFNEVIGRGHFGCVYHGTLLDNDGKKIHCAVKSLNRITDIGEVSQFLTEGIIMKDFSHPNVLSLLGICLRSEGSPLVVLPYMKHGDLRNFIRNETHNPTVKDLIGFGLQVAKGMKYLASKKFVH
RDLAARNCMLDEKFTVKVADFGLARDMYDKEYYSVHNKTGAKLPVKWMALESLQTQKFTTKSDVWSFGVLLWELMTRGAPPYPDVNTFDITVYLLQGRRLLQPEYCPDPLYEVMLKCWHPKAEMRPSFSELVSRISAIFSTFI
Mutation
P G Y F W V L I A T S Q N M C E D R K H
1059 1064 1069 1074 1079 1084 1089 1094 1099 1104 1109 1114 1119 1124 1129 1134 1139 1144 1149 1154 1159 1164 1169 1174 1179 1184 1189 1194 1199
1203 1208 1213 1218 1223 1228 1233 1238 1243 1248 1253 1258 1263 1268 1273 1278 1283 1288 1293 1298 1303 1308 1313 1318 1323 1328 1333 1338 1343
Position
l
4
3
2
1
0


B
Sequence
NPELVQAVQHVVIGPSSLIVHFNEVIGRGHFGCVYHGTLLDNDGKKIHCAVKSLNRITDIGEVSQFLTEGIIMKDFSHPNVLSLLGICLRSEGSPLVVLPYMKHGDLRNFIRNETHNPTVKDLIGFGLQVAKGMKYLASKKFVH
RDLAARNCMLDEKFTVKVADFGLARDMYDKEYYSVHNKTGAKLPVKWMALESLQTQKFTTKSDVWSFGVLLWELMTRGAPPYPDVNTFDITVYLLQGRRLLQPEYCPDPLYEVMLKCWHPKAEMRPSFSELVSRISAIFSTFI
Mutation
P G Y F W V L I A T S Q N M C E D R K H
1059 1064 1069 1074 1079 1084 1089 1094 1099 1104 1109 1114 1119 1124 1129 1134 1139 1144 1149 1154 1159 1164 1169 1174 1179 1184 1189 1194 1199
1203 1208 1213 1218 1223 1228 1233 1238 1243 1248 1253 1258 1263 1268 1273 1278 1283 1288 1293 1298 1303 1308 1313 1318 1323 1328 1333 1338 1343
Position
ll
3
2
1
0

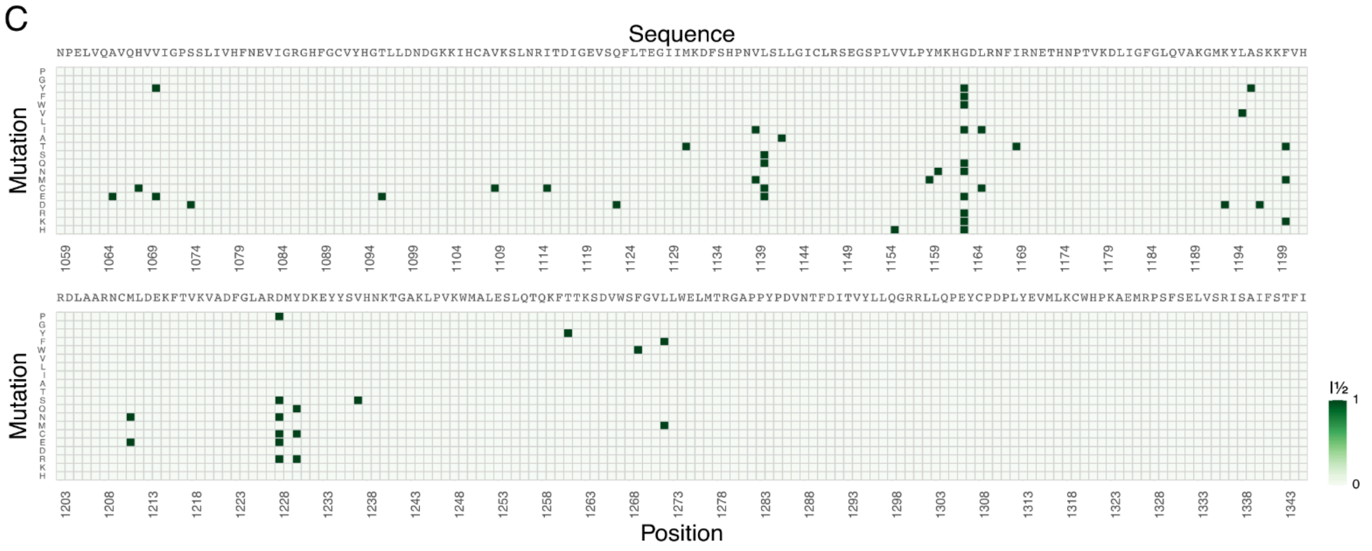

C
Sequence
NPELVQAVQHVVIGPSSLIVHFNEVIGRGHFGCVYHGTLLDNDGKKIHCAVKSLNRITDIGEVSQFLTEGIIMKDFSHPNVLSLLGICLRSEGSPLVVLPYMKHGDLRNFIRNETHNPTVKDLIGFGLQVAKGMKYLASKKFVH
RDLAARNCMLDEKFTVKVADFGLARDMYDKEYYSVHNKTGAKLPVKWMALESLQTQKFTTKSDVWSFGVLLWELMTRGAPPYPDVNTFDITVYLLQGRRLLQPEYCPDPLYEVMLKCWHPKAEMRPSFSELVSRISAIFSTFI
Mutation
P G Y F W V L I A T S Q N M C E D R K H
1059 1064 1069 1074 1079 1084 1089 1094 1099 1104 1109 1114 1119 1124 1129 1134 1139 1144 1149 1154 1159 1164 1169 1174 1179 1184 1189 1194 1199
1203 1208 1213 1218 1223 1228 1233 1238 1243 1248 1253 1258 1263 1268 1273 1278 1283 1288 1293 1298 1303 1308 1313 1318 1323 1328 1333 1338 1343
Position
l½
1
0

Figure 5 - figure supplement 1. Statistically filtered resistance mutations for grouped type I, type II, and type I½ inhibitors for MET. (A-C) Heatmaps of the sum of resistance mutations grouped for type I (pink), type II (blue), and type I½ (green) for MET.

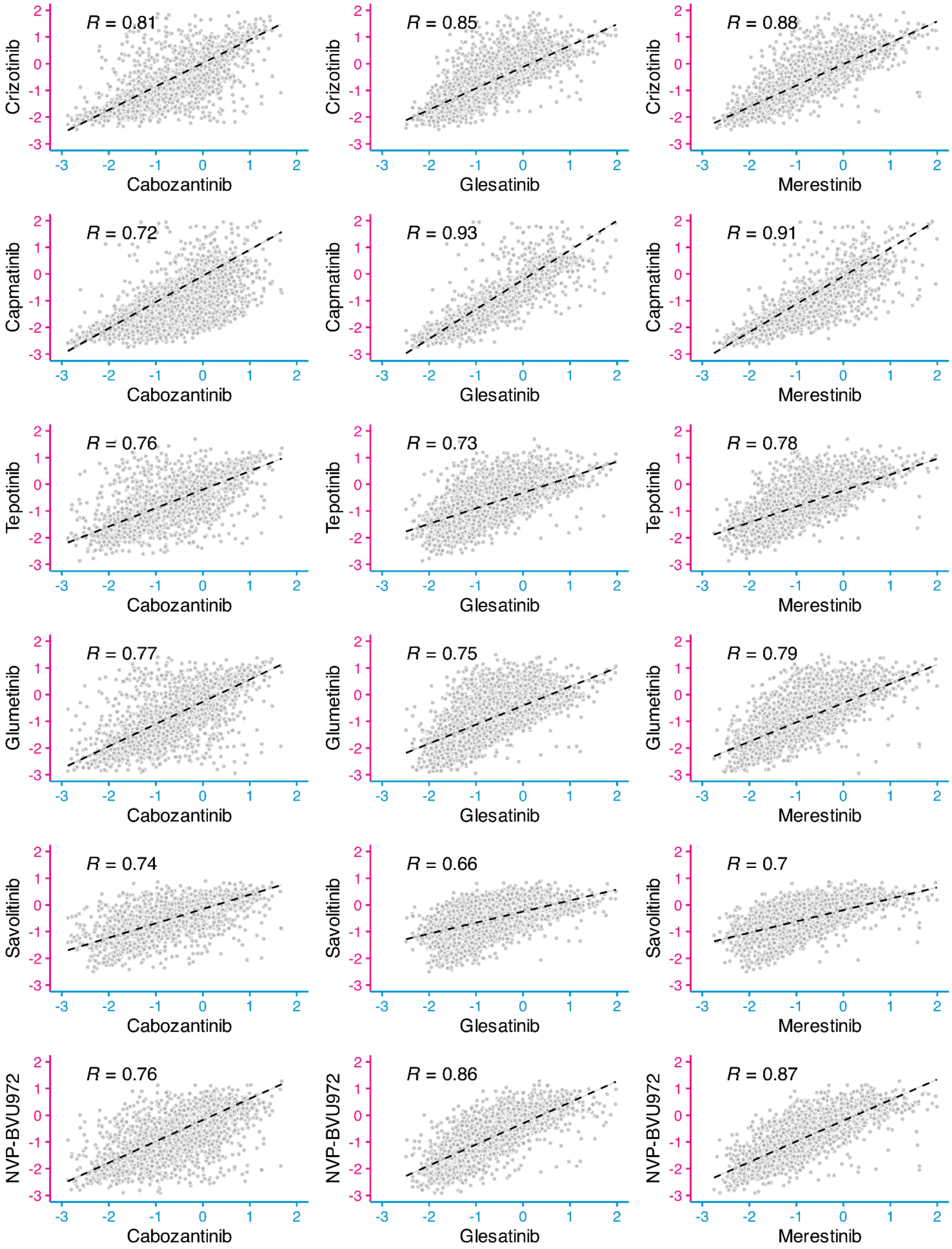

R = 0.81
R = 0.85
R = 0.88
Crizotinib
R = 0.72
R = 0.93
R = 0.91
Capmatinib
R = 0.76
R = 0.73
R = 0.78
Tepotinib
R = 0.77
R = 0.75
R = 0.79
Glumetinib
R = 0.74
R = 0.66
R = 0.7
Savolitinib
R = 0.76
R = 0.86
R = 0.87
NVP-BVU972
Cabozantinib
Glesatinib
Merestinib
-3
-2
-1
0
1
2

Figure 6 - figure supplement 1. Cross-comparison of type I and type II inhibitor pairs. Scatter plots of each type II inhibitor fitness scores (cabozantinib, glesatinib analog, merestinib; axis in blue) against each type I inhibitor (crizotinib, capmatinib, tepotinib, glumetinib, savolitinib, NVP-BVU972; axis in pink).

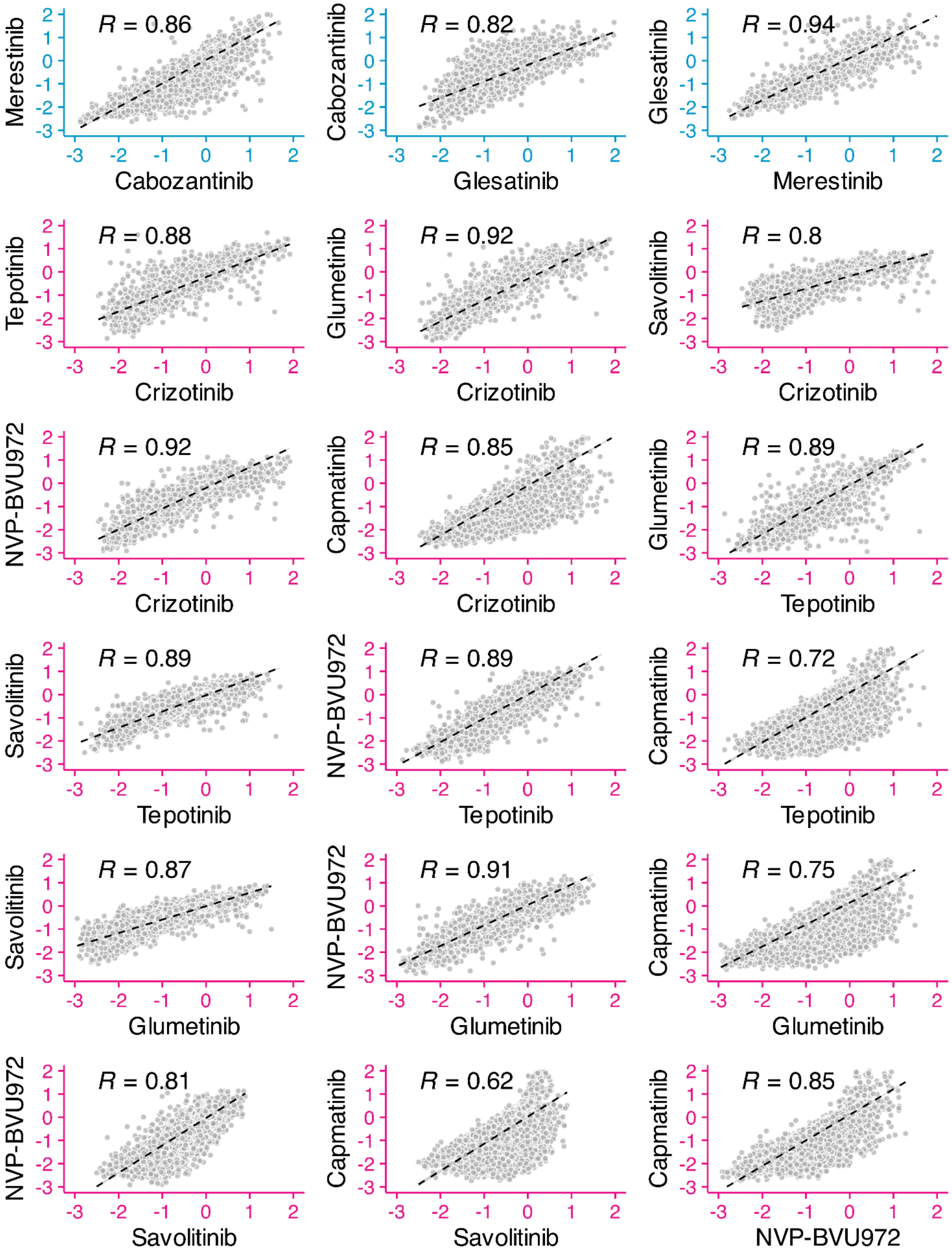


Figure 6 - figure supplement 2. Cross-comparison analysis of inhibitors within the same type. Scatter plots of each inhibitor pair within the type II group (cabozantinib, glesatinib analog, merestinib; axes in blue) and within the type I group (crizotinib, capmatinib, tepotinib, glumetinib, savolitinib, NVP-BVU972; axes in pink).

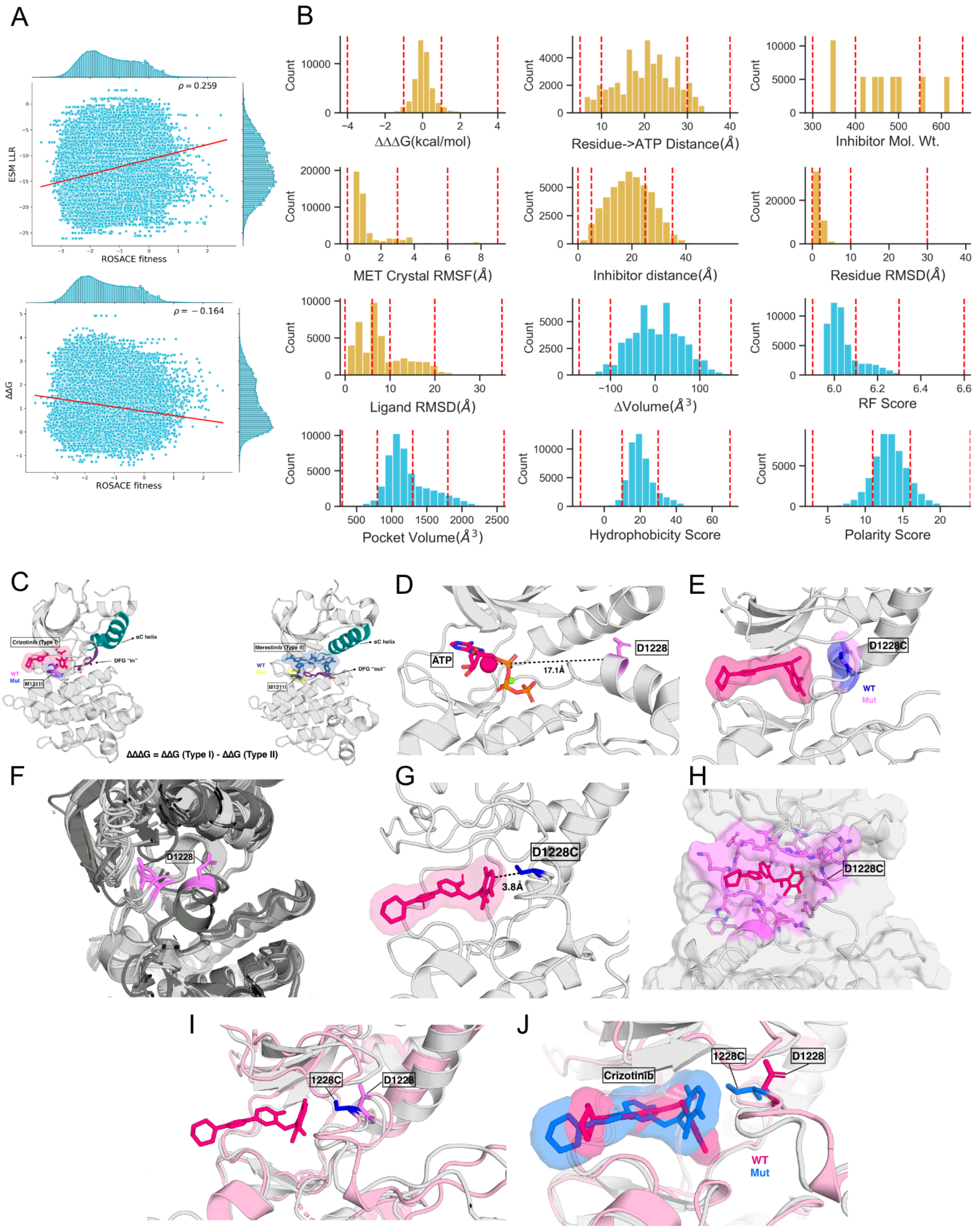

A
ρ = 0.259
ESM LLR
ROSACE fitness
ρ = −0.164
ΔΔG
ROSACE fitness
B
Count
ΔΔΔG(kcal/mol)
Residue->ATP Distance(Å)
Inhibitor Mol. Wt.
MET Crystal RMSF(Å)
Inhibitor distance(Å)
Residue RMSD(Å)
Ligand RMSD(Å)
ΔVolume(Å^3)
RF Score
Pocket Volume(Å^3)
Hydrophobicity Score
Polarity Score
C
Crizotinib (Type I)
αC helix
DFG "in"
WT
Mut
M1211I
Merestinib (Type II)
αC helix
DFG "out"
WT
Mut
M1211I
ΔΔΔG = ΔΔG (Type I) - ΔΔG (Type II)
D
ATP
17.1Å
D1228
E
D1228C
WT
Mut
F
D1228
G
D1228C
3.8Å
H
D1228C
I
1228C
D1228
J
Crizotinib
1228C
D1228
WT
Mut

Figure 7 - figure supplement 1. Distribution and visualization of features used in the XGBoost machine learning models. (A) Distribution of ESM LLR vs. Experimental fitness (top) and ΔΔG vs. Experimental fitness (bottom). (B) Distribution of all features (except ESM LLR and ΔΔG) extracted and used for the XGBoost models. The features that were incorporated in the best performing model are shown in yellow. The red dashed lines within each distribution show the edges of bins used to bin the feature values. (C) ΔΔΔG calculated from predicted ΔΔG Type I (PDB 2WGJ) (left) and Type II (PDB 4EEV) (right) MET kinase structure by subtracting type II ΔΔG from type I ΔΔG. The key regions showing difference in conformation between type I and II structures are the DFG motif (purple) and aC helix (teal). (D) Calculation of "residue to ATP" distance feature for residue D1228 in ATP bound MET Kinase structure (3DKC) is shown. Centroid of the ATP molecule is shown as a pink sphere. (E) Example of ΔVolume feature calculation using the difference between the volume of Asp and Cys. (F) Ensemble of MET kinase domain crystal structures aligned and RMSF of a given residue (D1228 in this example). (G) The shortest distance between the inhibitor and a mutation calculated from the Umol predicted variant-inhibitor structure. (H) The binding pocket of crizotinib in the predicted Umol structure. Pocket volume, hydrophobicity score, polarity score and RF score are calculated from this binding site. (I) Residue RMSD feature is described by the Umol predicted structure of variant D1228C (pink) superposed onto the wild-type reference structure (PDB 2WGJ, gray) and RMSD between D1228 in the wild-type structure and 1228C in the variant structure. (J) Ligand RMSD feature The Umol predicted structure of variant D1228C (pink) superposed onto the wild-type reference structure (PDB 2WGJ, gray) and RMSD between crizotinib in the wild-type structure (pink) and in the variant structure (blue).