# Supplementary Information

Computational Analysis of Kinase Inhibitor Selectivity using Structural Knowledge

Yu-Chen Lo1, Tianyun Liu1, 2, Kari M. Morrissey3, Satoko Kakiuchi-Kiyota4, Adam R. Johnson5, Fabio Broccatelli6, Yu Zhong4, Amita Joshi3, Russ B. Altman1,2

1Department of Bioengineering, Stanford University, Stanford, CA, USA

1. 2Department of Genetics, Stanford University, Stanford, CA, USA
2. 3Clinical Pharmacology, Genentech Inc., South San Francisco, CA, USA
3. 4Safety Assessment, Genentech Inc., South San Francisco, CA, USA
4. 5Biochemical and Cellular Pharmacology, Genentech Inc., South San Francisco, CA, USA
5. 6Drug Metabolism an Pharmacokinetics, Genentech Inc., South San Francisco, CA, USA

**\*Correspondence to:**

Russ B. Altman

Stanford Department of Bioengineering

Stanford, CA 90095

Phone: 650-725-0259

[Russ.Altman@stanford.edu](mailto:Russ.Altman@stanford.edu)

# Table of Contents

Suppl. Text 1. KinomeFEATURE database creation. ……..S4

Suppl. Text 2. Validation of KinomeFEATURE database using MEK kinase inhibitors ……..S4

Suppl. Text 3. Selectivity prediction performance assessment for 15 known kinase inhibitors ……..S5

Suppl. Text 4. Predicted kinase selectivity profile validation for 11 kinase drug candidates ……..S6

Suppl. Text 5. Structural alignment between intended and off-target structures from PDB search.....S7

Suppl. Fig. 1. MEK kinase inhibitor target profiling using the KinomeFEATURE database. ……..S9

Suppl. Fig. 2. Linear regression analysis between PFS and logKd of 15 kinase inhibitors. . ……..S10

Suppl. Fig. 3. Cross-similarity map of 55 kinase ATP binding sites ……..S11

Suppl. Fig. 4. Similarity threshold determination for the cross-similarity map of 55 kinases ...…….S12

Suppl. Fig. 5. ROC comparison between PocketFEATURE and sequence similarity search ……..S13

Suppl. Fig. 6. Dose-dependent titration curves of C16 and LY2874455……………………...….…..S14

Suppl. Fig. 7. Selectivity validation of 3 kinase compounds at 50% end-point inhibition……….…..S15

Suppl. Fig. 8. Effects of co-crystal structure differences on the PocketFEATURE scores…………..S16

Suppl. Fig. 9. Comparison of predicted and experimental IC50 values…..…………………….……..S17

Suppl. Fig. 10. PFS-IC50 linear correlation analysis for kinase inhibitor candidates **18-23** ….……...S18

Suppl. Fig. 11. PFS-IC50 linear correlation analysis for kinase inhibitor candidates **24-28**……...…..S19

Suppl. Fig. 12. Sequence alignment between primary and off-target structures from PDB search.....S20

Suppl. Fig. 13. Structural alignment between cMET and MER kinase structures ……..S21

Suppl. Table 1. Kinase structure distribution of the KinomeFEATURE database…….……..……..S22

Suppl. Table 2. PocketFEATURE performance assessment using 15 known kinase inhibitors….....S22

Suppl. Table 3. Performance comparison between the PocketFEATURE and the BSS approach….S22

Suppl. Table 4. Pair-wise PocketFEATURE scores of 55 kinase ATP sites.......................................S22

Suppl. Table 5. Pair-wise sequence identities of 55 kinase ATP sites……………....……………...S22

Suppl. Table 6. Selected list of inhibitors for17 kinase targets and their cross-activities against 55 kinases………...…………………………………………………………………………………......S22

Suppl. Table 7. Primary target prediction of 11 kinase inhibitors using the HNR and the KinomeFEATURE database………………………………………………………………………...S22

Suppl. Table 8. Target profiling of 11 kinase inhibitors using the KinomeFEATURE database.......S22

Suppl. Table 9. Experimental kinase activity profiling of 11 kinase inhibitors……………...….......S22

Suppl. References…………….…………….…………………………….........................................S22

**Supplementary Text 1. KinomeFEATURE database creation.**

To create the kinome database for the PocketFEATURE analysis, 292 kinase gene names from the SelectScreen® kinase selectivity panel (ThermoFisher) were converted to the corresponding UniProt ID using the retrieve/mapping protocol from the UniProt website (<http://www.uniprot.org/>). The mapping identified 283 UniProt IDs with the reviewed human protein sequences. To retrieve the corresponding kinase protein structures, the protein sequences of 283 UniProt IDs were compared against the PDB database sequences using the BLAST algorithm and homologous structures with > 50% identities were retained. Further structural filtering resulted in a total of 4582 PDB structures with at least one co-crystalized ligand. By comparing the UniProt IDs with those retrieved from the PDB, we yielded 2857 kinase structures that were identical to the 189 unique kinase genes in the kinase list (see Supplementary Table 1).

**Supplementary Text 2. Validation of KinomeFEATURE database using MEK kinase inhibitors.**

As a preliminary validation, we tested the capability of the KinomeFEATURE database to predict the primary target for an allosteric MEK kinase inhibitor, whose binding site is located adjacent to the ATP ligand (PDB ID: 4LMN) (Supplementary Fig. 1a). Structural alignment of the hit identified from the KinomeFEATURE database showed that the PocketFEATURE algorithm correctly identified MEK1 kinase structure from the top hits ranked by the PFS (PFS=-12) (Supplementary Fig. 1b and 1c). Likewise, a known off-target MEK2 was also correctly predicted (PFS=-9.4) (Supplementary Fig. 1b). To evaluate the advantage of the KinomeFEATURE database over a diversity protein database, we separately screened the MEK kinase pocket against the non-redundant protein database of 2000 human proteins based on the protein sequence diversity (<70%). While MEK1 and MEK2 targets were also predicted using the non-redundant protein database, their PFS were significantly higher (*i.e.*, less negative; -4.885 and -4.71, respectively) and would likely be misclassified as false negatives (Supplementary Fig. 1b). Furthermore, structural alignment based on microenvironment matching showed that the hit retrieved from the non-redundant database only achieved a partial overlap with the query (Supplementary Fig. 1d). In contrast, the identical structure was retrieved from the KinomeFEATURE database with full microenvironment overlaps (Supplementary Fig. 1d).

**Supplementary Text 3. Selectivity prediction performance assessment for 15 known kinase inhibitors.**

To quantitatively compare the computational predictions with experimental data for the 15 known kinase inhibitors, we further classified compounds as active or inactive at three binding affinity thresholds: 100nM (log*K*d=2), 1µM (log*K*d=3), 10µM (log*K*d=4) and assessed the prediction performance using different PocketFEATURE score cutoffs: -3, -4, and -5 (Supplementary Table 2). Next, we evaluated the accuracy, sensitivity and specificity of each inhibitor using the formula below, where TP=active and < PFS cutoff, TN=inactive and > PFS cutoff, FP=inactive and < PFS cutoff, FN=active and > PFS cutoff:

Accuracy =

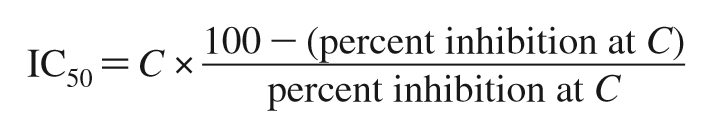
Sensitivity =

Specificity =

The average prediction accuracy at 100nM, 1µM and 10µM binding affinity thresholds was 91%, 85%, and 78%, respectively, using the optimal scoring cutoff (PFS=-5 or lower) (Supplementary Table 2). Furthermore, the average specificity evaluated at all binding affinity levels was >95% (Supplementary Table 2). In a retrospective study utilizing the same data set, we compared the performance of PocketFEATURE with the binding site signature (BSS) approach by estimating the binding site similarity based on the interaction energy between ligand and critical residues(Subramanian and Sud, 2010). Using the optimal PFS cutoff, the PocketFEATURE algorithm achieved a higher average accuracy (90% vs 87%) and specificity (96% vs 89%) than the BSS method at the 100-nM potency threshold (Supplementary Table 3)(Karaman, et al., 2008). On the other hand, the average sensitivity is slightly lower than the BSS approach (40% vs 45%).

**Supplementary Text 4. Predicted kinase selectivity profile validation for 11 kinase drug candidates**

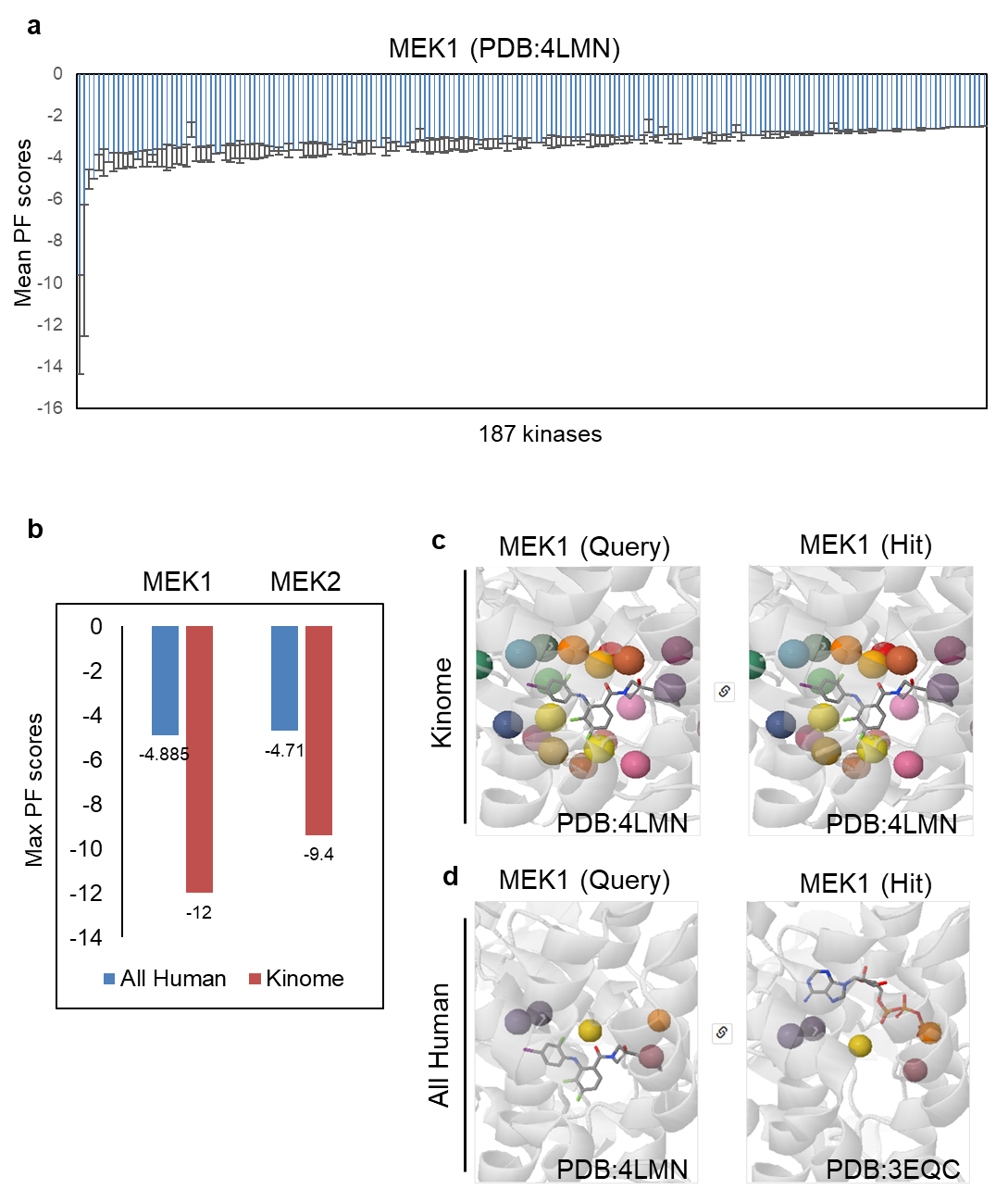
We validated the predicted off-target binding of these compounds by quantifying their inhibition of kinase activity or binding at 0.1 µM or 1 µM against 111 human kinase assays in the SelectScreen® kinase profiling panel (Supplementary Table 9). Initial selectivity validation of 3 kinase inhibitors: JAK inhibitor (compounds **21**) at 1 µM concentration, MET inhibitor (compound **24**) at 0.1 µM concentration and B-RAF inhibitor (compound **26**) at 1 µM concentration using a 50% inhibition threshold showed that the PFS cutoff of -5 can selectively identify primary targets and other off-targets (Supplementary Fig. 7). To evaluate the impact of the co-crystallized structure on the selectivity prediction, we analyzed and compared homologous co-crystal structures of a JAK kinase inhibitor (compound **21**) with its predicted target binding profile. The correlation analysis of the PFS values between the homologous JAK co-crystal structures did not reveal significant differences in the PFSs (Supplementary Fig. 8) except for JAK1 and TYK2 kinases. To incorporate a concentration-dependent effect, we estimated IC50 values of 11 compounds based on their % inhibition at 1 µM by assuming the % inhibition data would adhere to a classical sigmoidal inhibition model(Sciabola, et al., 2008). Estimated kinase IC50s were calculated from the single point % inhibition assay data using the following equation(Sciabola, et al., 2008):



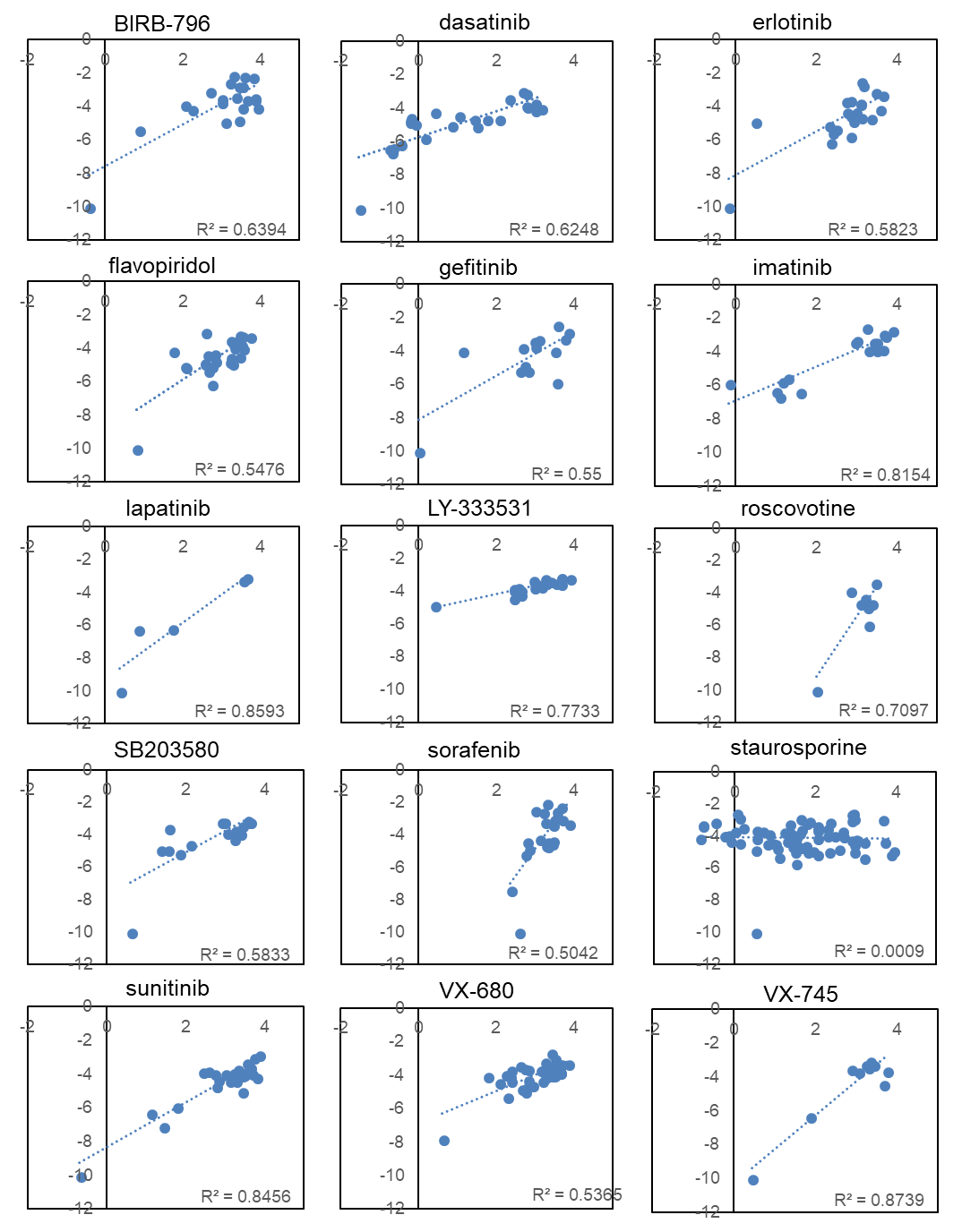
In the equation, the constant *C* represents the inhibitor test concentration. To validate this approach, we correlated the experimentally determined IC50 values of 85 kinase inhibitors with their predicted IC50 values as extrapolated from their % inhibition values at 1 µM (Supplementary Fig. 9). The linear regression analysis of predicted logIC50 and experimental logIC50 showed a high correlation with a slope of 0.9 and R2 > 0.8.

**Supplementary Text 5. Structural alignment between intended and off-target structures from PDB search.**

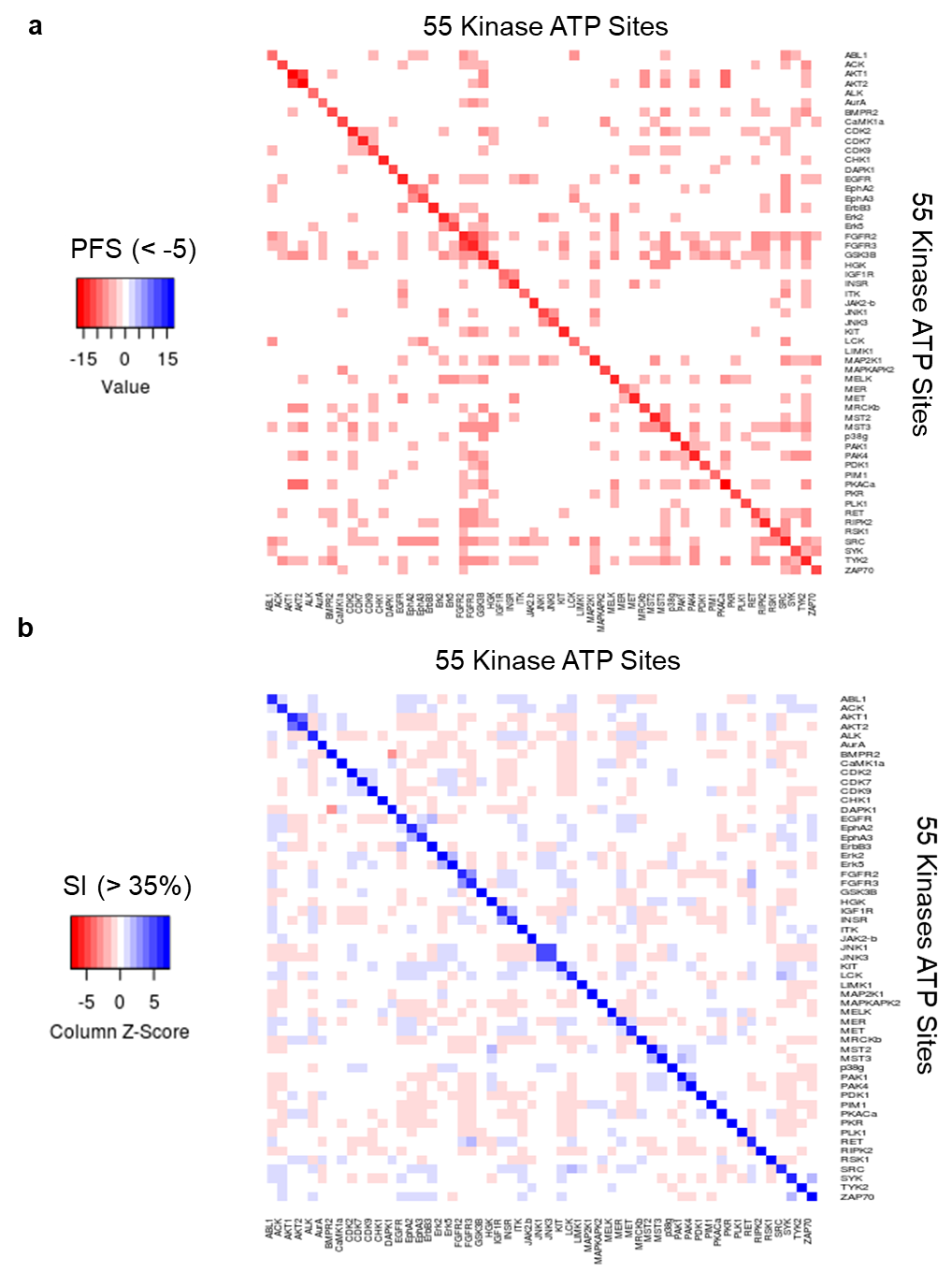
To explain the cross-activity mechanism of PIM (compound **18**), JAK (compound **22**) and MET (compound **24**) kinase drug candidates, we performed structural alignments between their intended targets: PIM (PDB: 1XR1), JAK (PDB: 4IVD) and MET (PDB: 3LQ8) and off-targets: GSK3β (PDB: 3PUP), TYK2 (PDB: 3LXN), and LCK (PDB: 2OG8) from the KinomeFEATURE database search. The structural alignment was performed using the MatchMaker protocol in the UCSF Chimera program that minimizes the structural overlap based on the root-mean-square deviation (RMSD) measurement (Supplementary Fig. 12a). In addition, CHK1 structure (PDB: 1NVR) with previously annotated ATP-competitive kinase pharmacophores, was also structurally aligned with the selected structures(Liao, 2007; Traxler and Furet, 1999). Based on the multiple sequence alignment (MSA) derived from the structural alignment, we identified critical residues that involved in several regions important for the ATP-competitive binding including the phosphate binding region/DFG, the ribose/adenine binding region, front specificity pocket, linker region, gate keeper residue, and hydrophobic pocket (Supplementary Fig. 12b). The phosphate binding region, including the aspartate of the DFG activation loop is important for the positioning of ATP phosphates while the linker region, ribose/adenine binding region makes critical contact with the ATP ribose and adenine moiety. Front specificity pocket is a relatively small hydrophobic region at the transition zone between ATP binding pocket and the hydrophilic solvent exposed area of the protein, which consists of variable hydrophilic residues that are useful for modulating kinase inhibitor specificity. The hydrophobic pocket, also known as “back hydrophobic pocket”, is located on the backend of the ATP ligand and the entry to this pocket is controlled by the gate keeper residue, such as L84 for CHK1 (see Supplementary Fig. 12) which often determines kinase inhibitor specificity as well as their resistance. The critical residues involved in these regions were then used to identify the corresponding microenvironment pairs and site-specific PFS values using the PocketFEATURE algorithm (Fig. 5c).

****

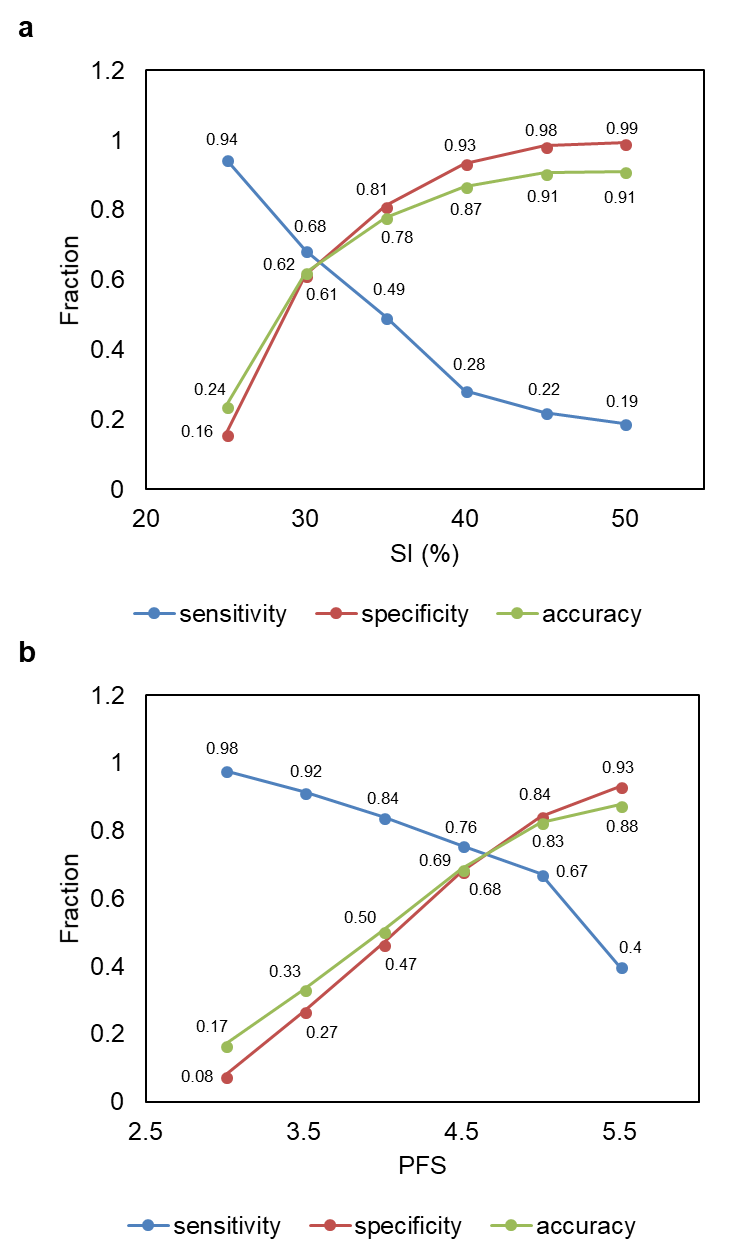
**Supplementary Fig. 1. MEK kinase inhibitor target profiling using the KinomeFEATURE database.** (**a**)The MEK kinase pocket (PDB: 4LMN) was compared to 3000 kinase pockets in the KinomeFEATURE database based on the microenvironment similarity using the PocketFEATURE algorithm. The average PFS of each of 187 unique kinases was evaluated and the most negative PFS was used to predict the on-target. (**b**) Comparison of the maximum PFS for on-target MEK1 and off-target MEK2 using the all human non-redundant database (All Human) or the KinomeFEATURE database (Kinome). Note that the PFS is more negative using the KinomeFEATURE database for both MEK1 and MEK2 kinases. (**c**) Structural alignment of the hit identified from the KinomeFEATURE database showed that the PocketFEATURE algorithm correctly identified MEK1 kinase structure from the top hits ranked by the PFS (PFS=-12). (**d**) Structural alignment based on microenvironments matching showed that the hit retrieved from the non-redundant human database only achieved a partial alignment with the query.

****

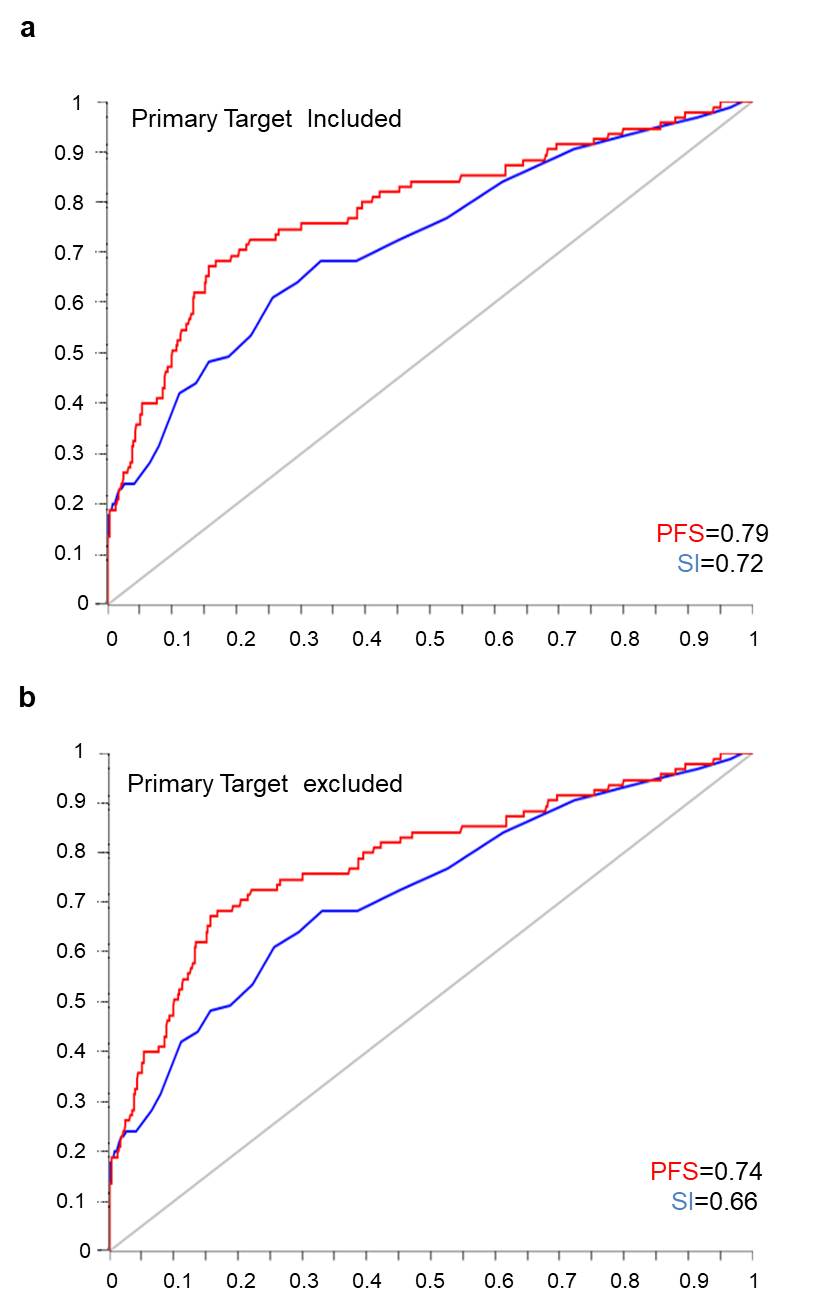
**Supplementary Fig. 2.** **Linear regression analysis between PFS and logKd of 15 kinase inhibitors.** With the exception of staurosporine, all inhibitors have an R2 value > 0.5. The average R2 values for the kinase inhibitors are 0.63 (including staurosporine) and 0.69 (excluding staurosporine).



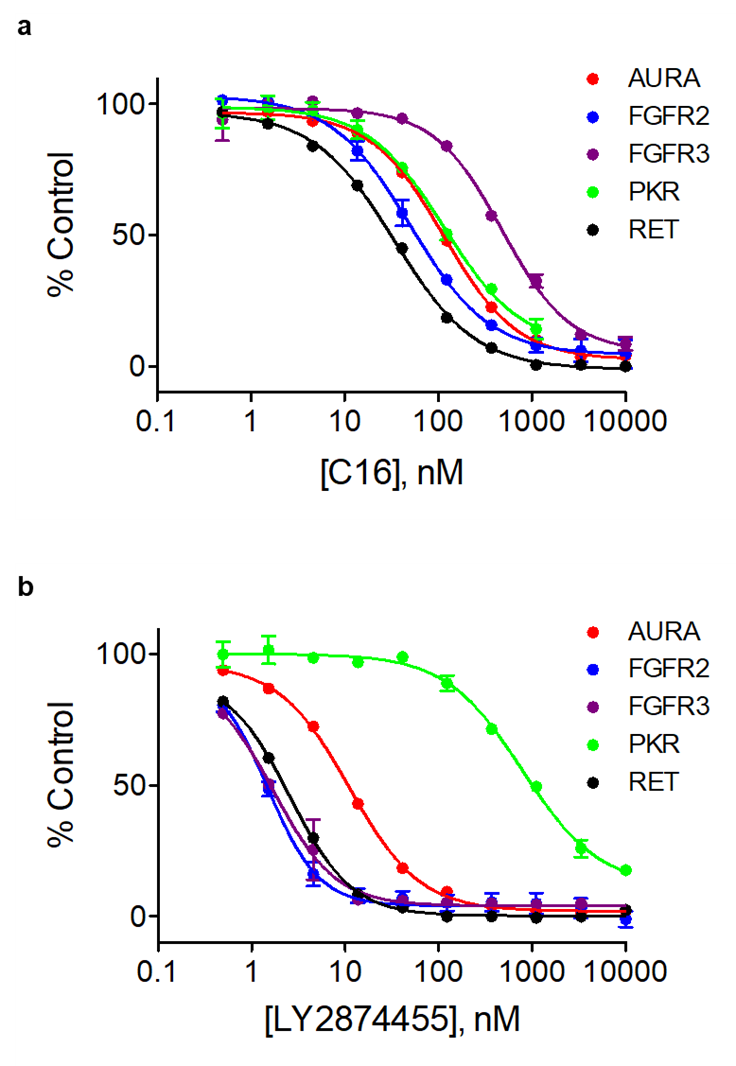
**Supplementary Fig. 3. Cross-similarity map of 55 kinase ATP binding sites.** Mapping of 55 pair-wise similarity scores of kinase ATP binding sites using (**a**) the microenvironment matching based on the PocketFEATURE scores (PFS<-5) or (**b**) the sequence homology search based on the sequence identity (SI> 35%). To facilitate comparison, the sequence identities in each column were converted to column Z-scores.

****

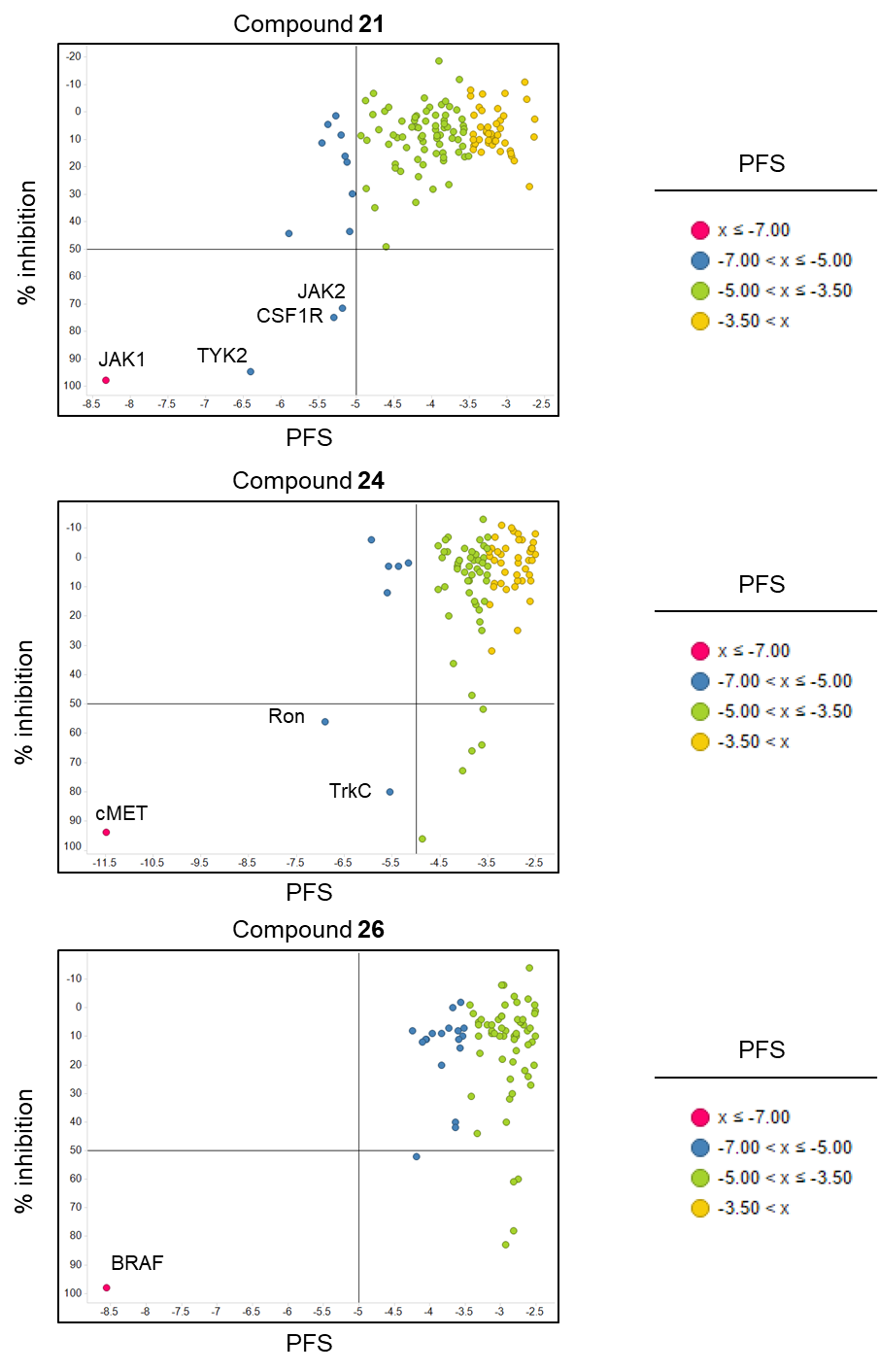
**Supplementary Fig. 4. Similarity threshold determination for the cross-similarity map of 55 kinases.** To identify the optimal similarity threshold based on (a) PFS or (b) SI, 17 kinases with at least one specific inhibitors were identified from a previous kinase selectivity panel (see Supplementary Table 6). The off-kinase binding data were used to determine the accuracy, specificity and sensitivity at multiple sequence identity (25-50%) or PFS values (-3 to -5.5). Note that absolute PFS values were used for comparison.

****

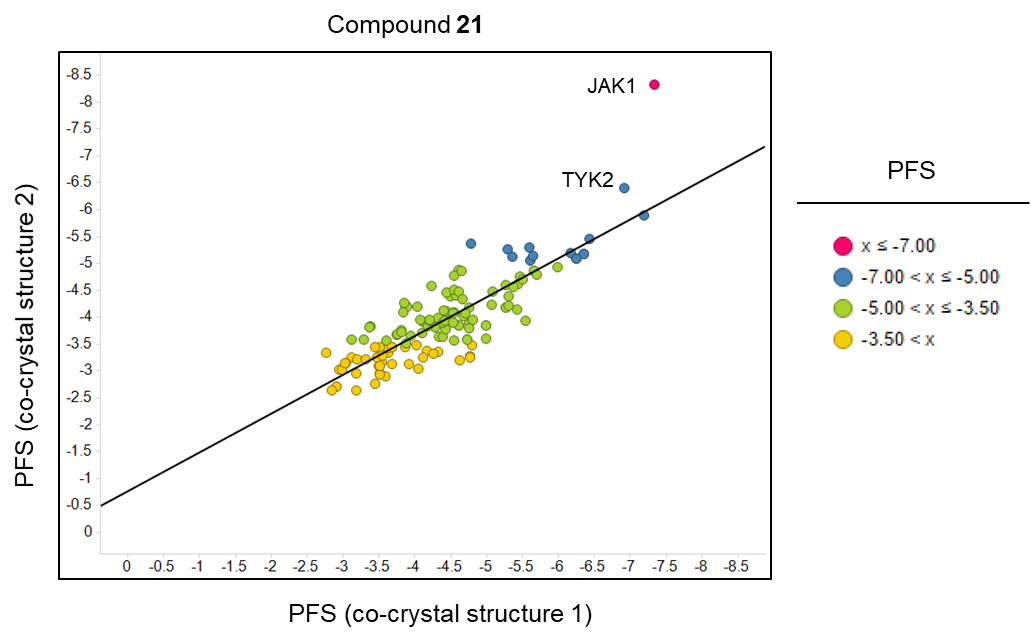
**Supplementary Fig. 5.**  **ROC comparison between PocketFEATURE and sequence similarity search.** AUC values evaluated based on the ROC curves showed that PFS has a higher AUC scores than SI by including primary targets (AUCPFS=0.79 vs AUCSI=0.72) and excluding primary targets (AUCPFS=0.74 vs AUCSI=0.66). The removal of primary targets did not significantly affect the performance using either approaches, suggesting that improved performance by PFS was not attributed to correct prediction of primary targets.



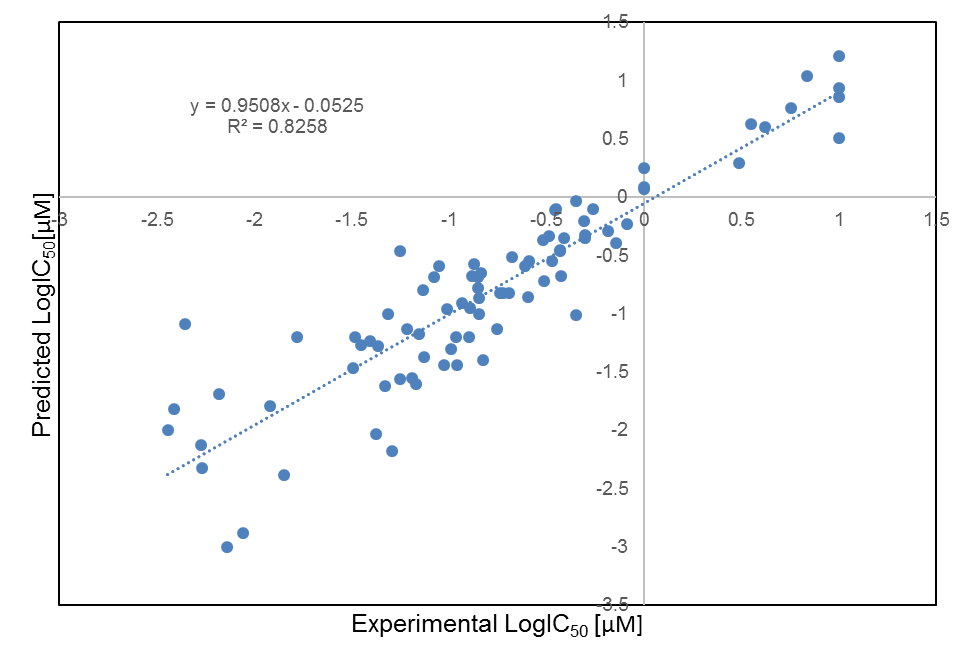
**Supplementary Fig. 6. Dose-dependent titration curves of C16 and LY2874455.** The kinase PKR specific inhibitor C16 (a) and FGFR specific inhibitor LY2874455 (b) were evaluated for their selectivity against 5 predicted kinases (AURA, FGFR2, FGFR3, PKR and RET) by dose-dependent titrations using a test concentration range from 1nM to 10µM. Note that the y-axis indicated the %inhibition relative to control. See Fig. 4c for compound potency (IC50).



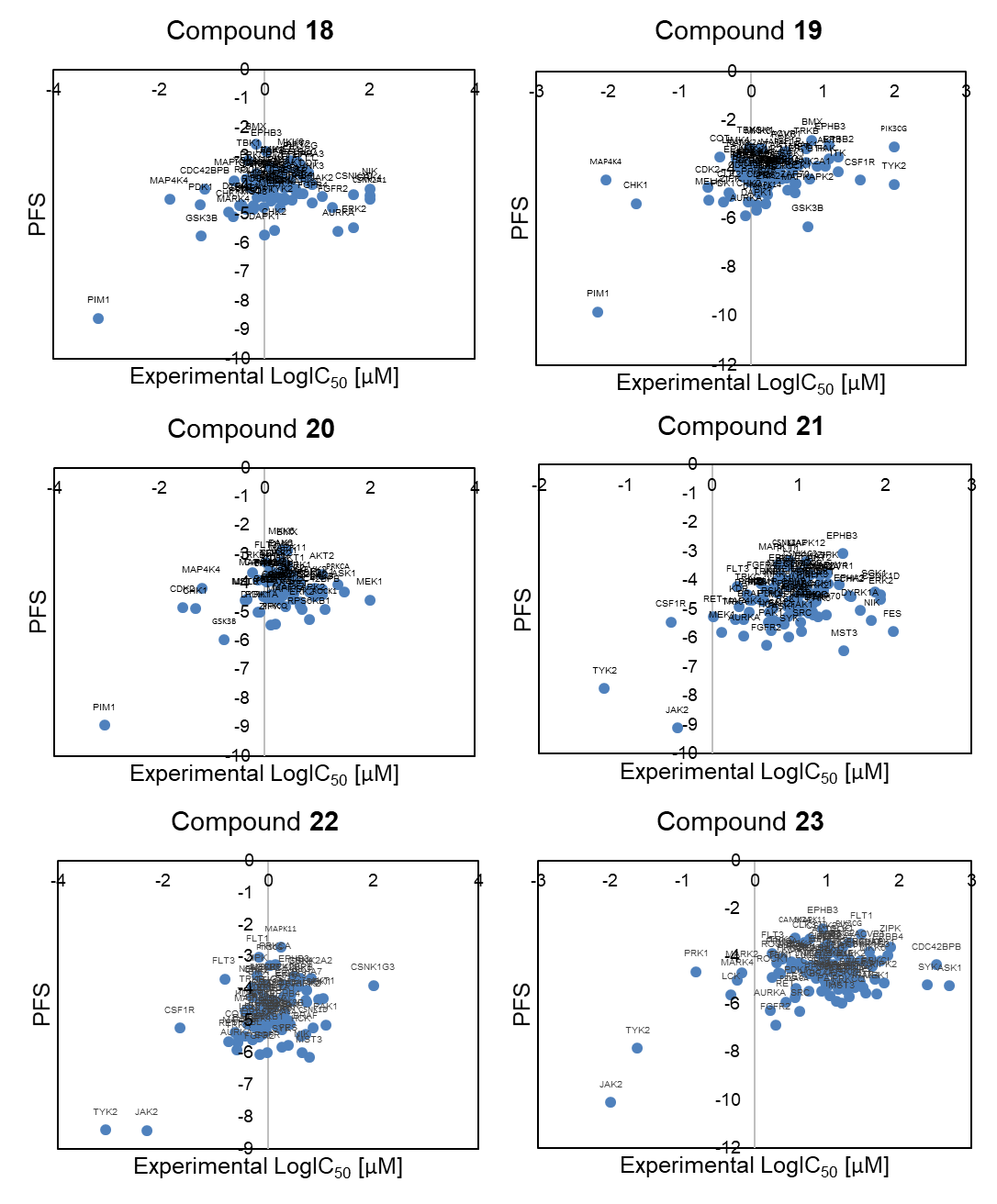
**Supplementary Fig. 7**. **Selectivity validation of 3 kinase compounds at 50% end-point inhibition.** Percentage (%) activity inhibition of JAK, cMET and BRAF kinase inhibitors (compound **21**, **24** and **26**) against the kinase panel were correlated with their PFS values. Note that the data points are kinases colored based on their PFS values.

****

**Supplementary Fig. 8**. **Effects of co-crystal structure differences on the PocketFEATURE scores.** Homologous co-crystal structures of a JAK kinase drug candidate (compounds **21**) were used for off-target profiling and the predicted PFS were compared with the original structures. Except for two targets, JAK1 and TYK2, no significant changes in PFS values were observed.

****

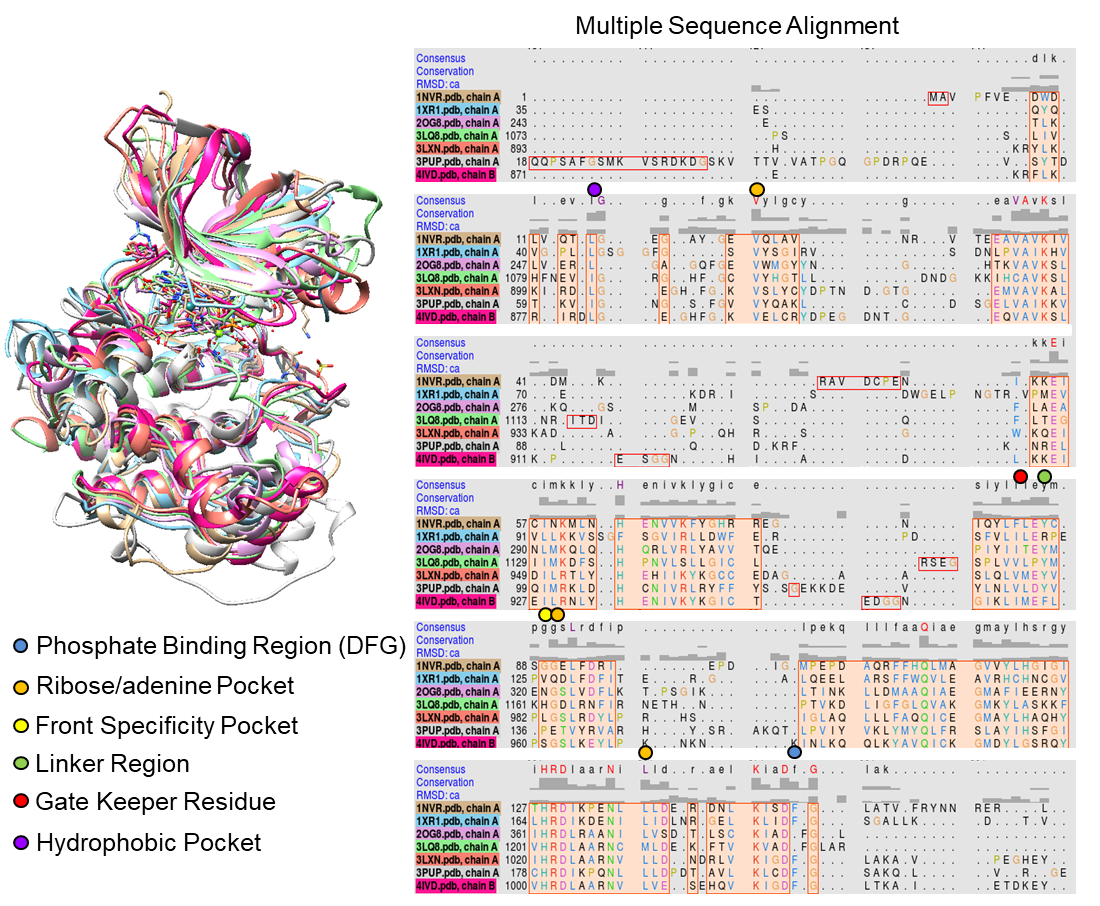
**Supplementary Fig. 9. Comparison of predicted and experimental IC50 values.** 65 predicted IC50 values predicted from the experimentally determined % inhibition at 1 µM were compared with the experimentally determined IC50s. The linear regression analysis of predicted logIC50 and experimental logIC50 shows a high correlation with a slope of 0.9 and R2 value > 0.8.

****

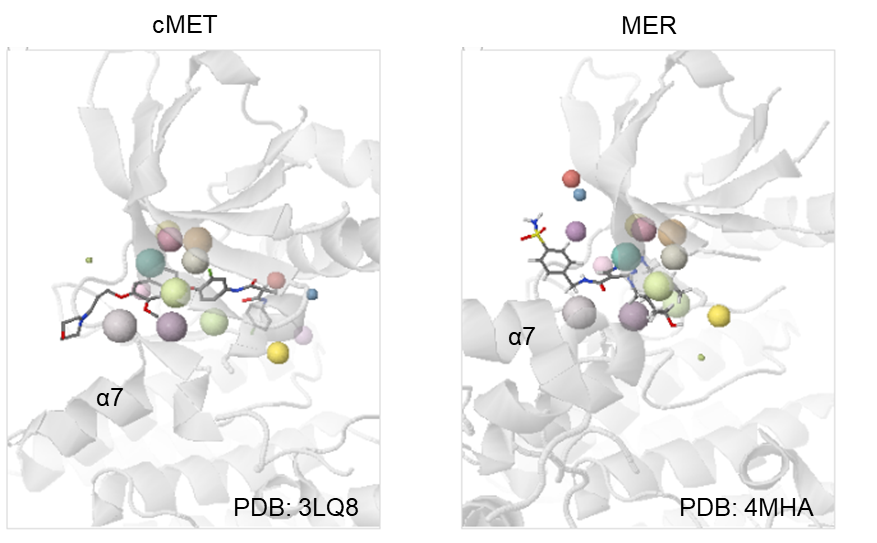
**Supplementary Fig. 10. PFS-IC50 linear correlation analysis for kinase inhibitor candidates 18-23.** The consensus off-targets for PIM inhibitors (compounds **18**, **19**,and **20**) are CHK1, DAPK1 and GSK3β. The consensus off-targets for JAK inhibitors (compounds **21**, **22**, and **23**) are TYK2 and CSF1R. Note that primary targets are identified by the most negative PFS.

****

**Supplementary Fig. 11. PFS-IC50 linear correlation analysis for kinase inhibitor candidates 24-28.** The consensus off-targets for MET inhibitors (compounds **24** and **25**) are TRK, LCK and FLT3. The consensus off-targets for BRAF inhibitors (compounds **26** and **27**) are LCK and CSF1R. On the other hand, the PI3K inhibitor (compound **28**) reveals low off-target promiscuity. Note that the primary targets are identified by the most negative PFS.

****

**Supplementary Fig. 12. Sequence alignment between primary and off-target structures from PDB search.** (**a**) Structural alignment between primary targets: PIM1 (PDB: 1XR1), JAK (PDB: 4IVD), and cMET (PDB: 3LQ8) and off-targets: GSK3β (PDB: 3PUP), TYK2 (PDB: 3LXN), and LCK (PDB: 2OG8) with CHK1 kinase (PDB: 1NVR) were used to derive multiple sequence alignment and identify critical residues participating in competitive ATP binding. The binding regions include the phosphate binding region (DFG) (color: blue), the ribose/adenine pocket (color: orange), the front specificity pocket (yellow), the linker region (color: green), the gate keeper residue (color: red), and the hydrophobic pocket (color: purple).

****

**Supplementary Fig. 13. Structural alignment between cMET and MER kinase structures.**

Although MER is an off-target of the cMET inhibitor (compound **24**) and its logIC50 value is comparable to that of cMET (logIC50, MER=-2.38 *vs*. logIC50, cMET=-2.18), its PFS value is significantly lower than that of cMET kinase (PFSMER=-5.55 vs PFScMET=-14.74) (Supplementary Tables 8 and 9). This deviation could be explained by the structural alignment between the predicted on-target structure, cMET (PDB: 3LQ8) and off-target structure, MER (PDB: 4MHA) where structural changes upon ligand binding involving helix α7 significantly reduce protein microenvironment overlaps between the two binding pockets.

**Supplementary Table 1. Kinase structure distribution of the KinomeFEATURE database.**

**Supplementary Table 2. PocketFEATURE performance assessment using 15 known kinase inhibitors.** The 15 known inhibitors are classified as active or inactive at three binding affinity thresholds: 100 nM (log*K*d=2), 1µM (log*K*d=3), 10µM (log*K*d=4) and assessed the prediction performance using different PFS cutoffs at -3, -4, and -5. Note that 6 protein structures (PDB: 2HYY, 3GVU, 1T46, 3HEC, 2PL0, and 1XBB) were used to evaluate the performance for imatinib inhibitor based on their average PFS values and binding affinity data.

**Supplementary Table 3. Performance comparison between the PocketFEATURE and the BSS approach.**

**Supplementary Table 4. Pair-wise PocketFEATURE scores of 55 kinase ATP sites.**

**Supplementary Table 5. Pair-wise sequence identities of 55 kinase ATP sites.**

**Supplementary Table 6. Selected list of 17 kinase inhibitors and their cross-activities against 55 kinases.** The compounds were previously reported from a large-scale kinase activity assay(Davis, et al., 2011).

**Supplementary Table 7. Primary target prediction for 11 kinase inhibitors using HNR and KinomeFEATURE.**

**Supplementary Table 8. Target profiling of 11 kinase inhibitors using the KinomeFEATURE database.** The top predicted PocketFEATURE scores for each kinase were mapped to 111 kinases in SelectScreen® kinase selectivity panel.

**Supplementary Table 9. Experimental kinase activity profiling of 11 kinase inhibitors.** The % inhibition was measured at 0.1 or 1 µM using the SelectScreen® kinase selectivity panel consisting of 111 kinases and then converted to calculated logIC50 values (see Supplementary Text 4).

**Supplementary References**

Davis, M.I.*, et al.* Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* 2011;29(11):1046-1051.

Karaman, M.W.*, et al.* A quantitative analysis of kinase inhibitor selectivity. *Nat Biotechnol* 2008;26(1):127-132.

Liao, J.J. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J Med Chem* 2007;50(3):409-424.

Sciabola, S.*, et al.* Predicting kinase selectivity profiles using Free-Wilson QSAR analysis. *J Chem Inf Model* 2008;48(9):1851-1867.

Subramanian, G. and Sud, M. Computational Modeling of Kinase Inhibitor Selectivity. *ACS Med Chem Lett* 2010;1(8):395-399.

Traxler, P. and Furet, P. Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. *Pharmacol Ther* 1999;82(2-3):195-206.