Objectives

We have previously demonstrated the importance of using CATH-FunFams as a reasonable annotation level for drug-domain interaction and also used network analysis to associate the propensity of side effects with the network properties of these families through our druggable FunFam approach [Moya-García et al., 2017]. In this study, we are focusing on protein kinase inhibitors, a set of molecules that inhibit the activities of kinases. Using a set of publicly available protein kinase inhibitors as well as FDA-approved kinase inhibitor drugs, we studied the Functional Families of kinases and associated these FunFams with protein kinase inhibitors. We subsequently measured some network characteristics to distinguish FunFams with side effects from others. AMG:FF don't have SEs per se. They might contain proteins that are associated with drug side effects. Check out Duran-Frigola, M. & Aloy, P. Analysis of chemical and biological features yields mechanistic insights into drug side. We also explored some similarity measures to shed more lights on possible repurposing of protein kinase inhibitors to members of a given FunFam. The outline of the study is summarized below

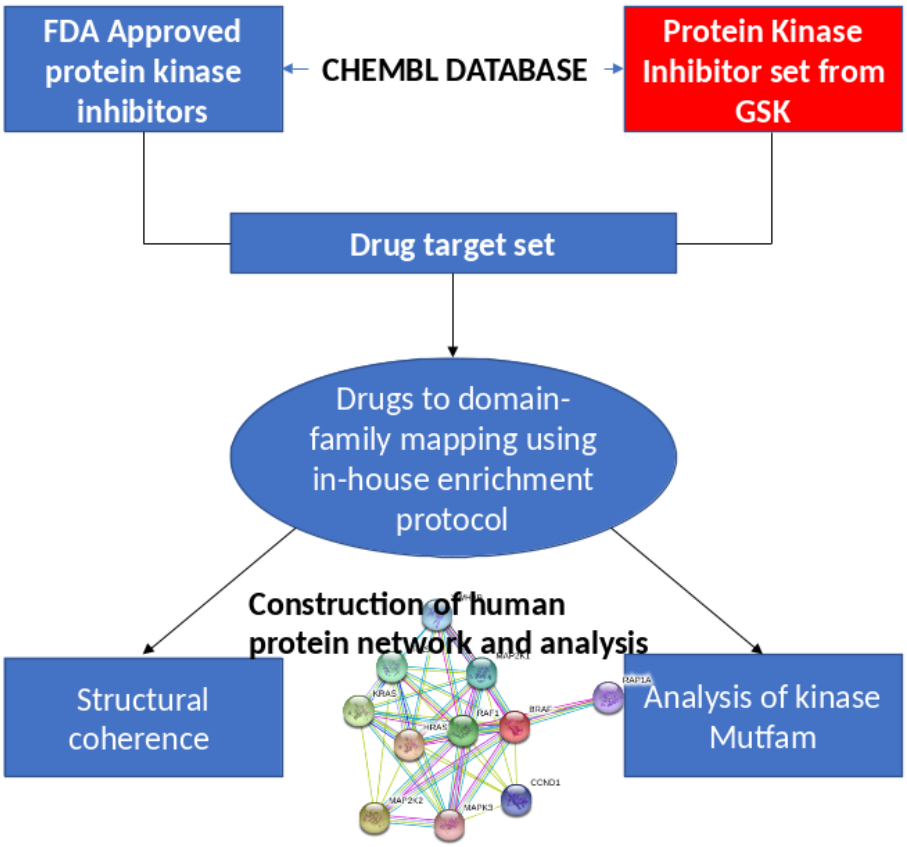


Figure 1: Frame-work of the methodology used in this study

Results and Discussion

Protein Kinase Inhibitors set

The Published Kinase Inhibitor Set (PKIS) is a collection of 367 compounds that have been made available by GSK to the ~~external~~ research community [Dranchak et al., 2013, Knapp et al., 2013]. These compounds have been annotated with protein kinase activity [Knapp et al., 2013] and are of various chemotypes and are openly available from the ChEMBL database (ChEMBL release 23) [Gaulton et al., 2016]. The PKIS are active against some known target kinases and can be extended to other new target kinases. We collected a subset of the PKIS that showed an inhibitory activity level above 50% as [Dranchak et al., 2013, Anastassiadis et al., 2011] had reported this threshold as appropriate for considering the inhibition of kinase catalytic activity. This gave a unique set of 205 protein kinase inhibitors that were distributed across 133 protein kinases. This covers about 60% of the entire PKIS set and is thus a reasonable dataset to consider for a network assessment and characterization of protein kinase inhibition.

Furthermore, we compiled kinase-inhibitor datasets of the FDA-approved drugs by querying ChEMBL release 23. We defined this drug target set based on the concentration with which the drug affects the protein. We considered a drug as a small molecule with therapeutic application, which has a direct binding to a single protein (ASSAY_TYPE ="B"), with a maximaum phase of development ="4" which indicate that the drug has been approved. We filtered out weak activity by considering drug-target activity stronger than 1μM and a pchembl_value$\geq$6). The ATC-code was used to filter the drug-targets such that only the protein kinase inhibitor was obtained. The ATC code classifies drug into different groups at different levels. The code "L01XE" correspond to antineoplastic drugs which are protein kinase inhibitors. The approved drug-target set are 29 approved drugs that interacts with 324 targets. The targets were filtered to exclude those that are not kinases reducing the numbers of targets to 305 kinases. This was used for further analysis.

GSK-Protein kinase inhibitors profile

We compiled the various dataset obtained from ChEMBL-23 that correspond to the GSK-Protein kinase inhibitors.

[this sentence does not contribute any meaning or additional info]

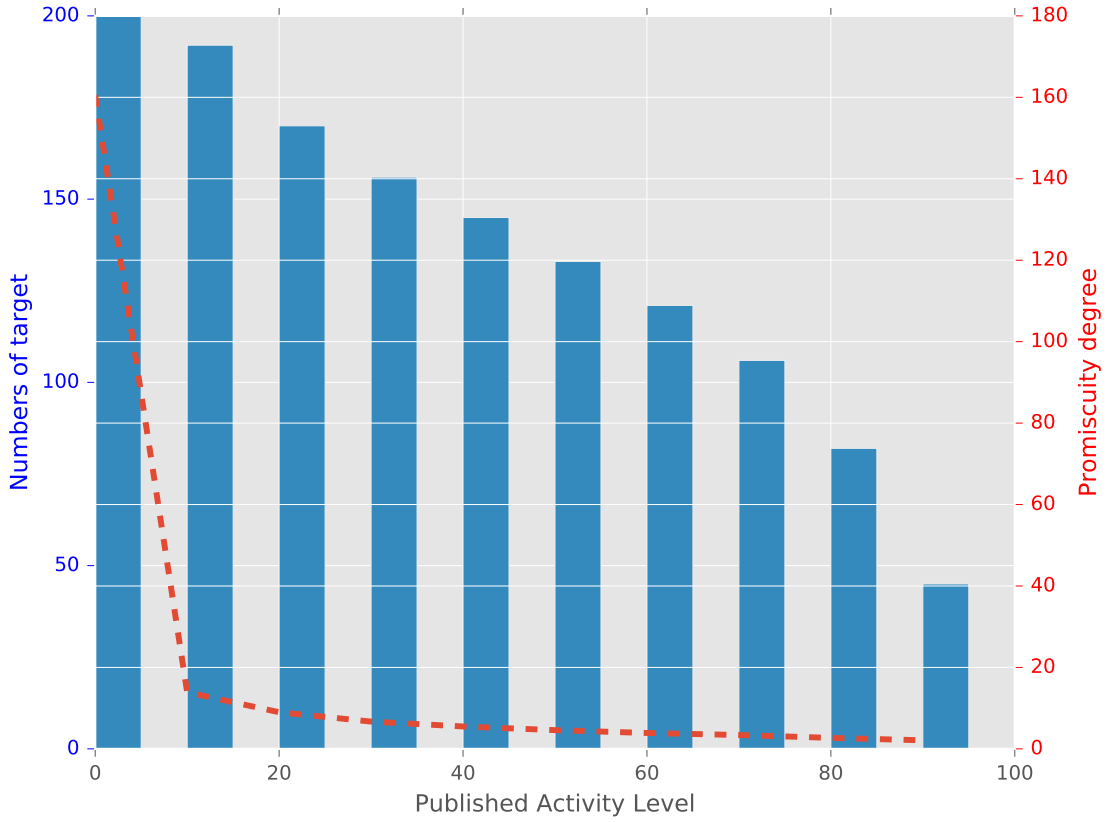


Figure 2: Promiscuity profile of the dataset used in this study. The promiscuity degree correspond to the number of possible interactions per drug in the dataset. The bar plot hence shows a drop of over 10 folds for the promiscuity degree by 10% increase in the activity threshold (0-10). This significant drop in the promiscuity degree was not observed for the other activity threshold. Hence, we choose those dataset with an activity level $\geq$50%.

[I don't understand this plot. What are the "possible interactions"?]

Comparison of the FDA drug-target set and the GSK-PKI drug-target

We had two different drug target sets. Both drug-target sets were selected using different criteria as they were created using different approach. Hence, we compared the network properties of the targets from both sets. The targets from the GSK-PKIS is a subset of the FDA-approved PKI targets, however, the drugs are distinct and not shared in both dataset. [This doesn't make sense: 1 maybe the smaller FDA data is a subset of the larger GSK data. 2 If the drugs are distinct between the two sets, one set cannot be a subset of the other] Using the [Menche et al., 2015] score called "DS-Score" [This paragraph should be rewritten, also it is not a "given" network is the protein functional network you modelled after STRING data]

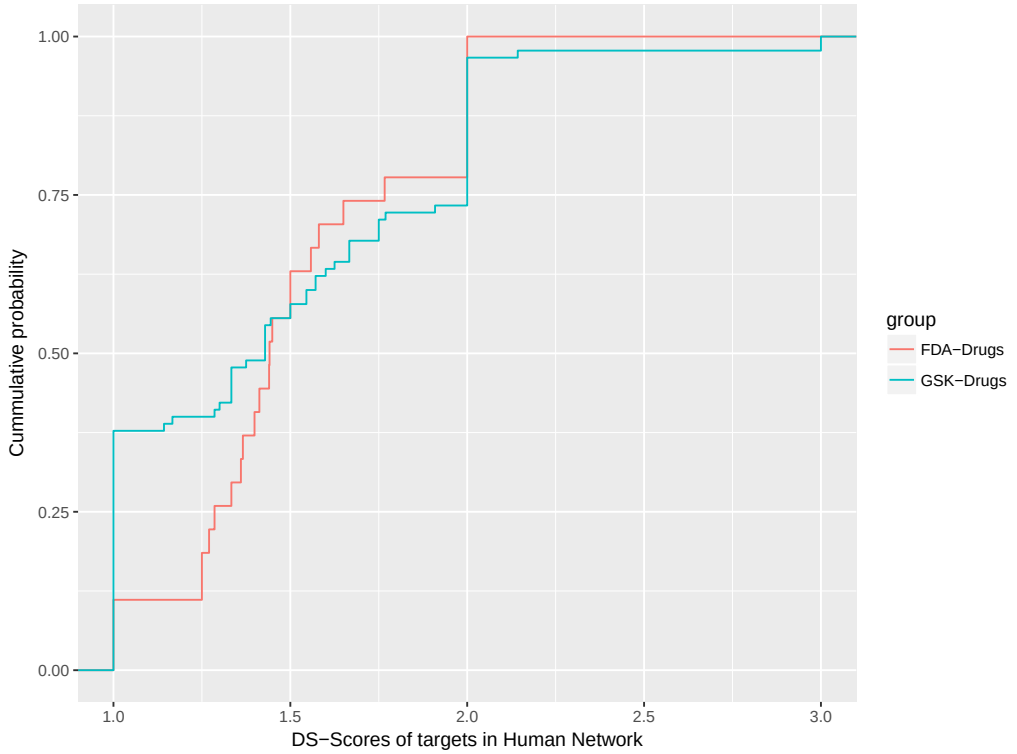


Figure 3: Comparing the similarity measure of targets of FDA and GSK-PKIS sets

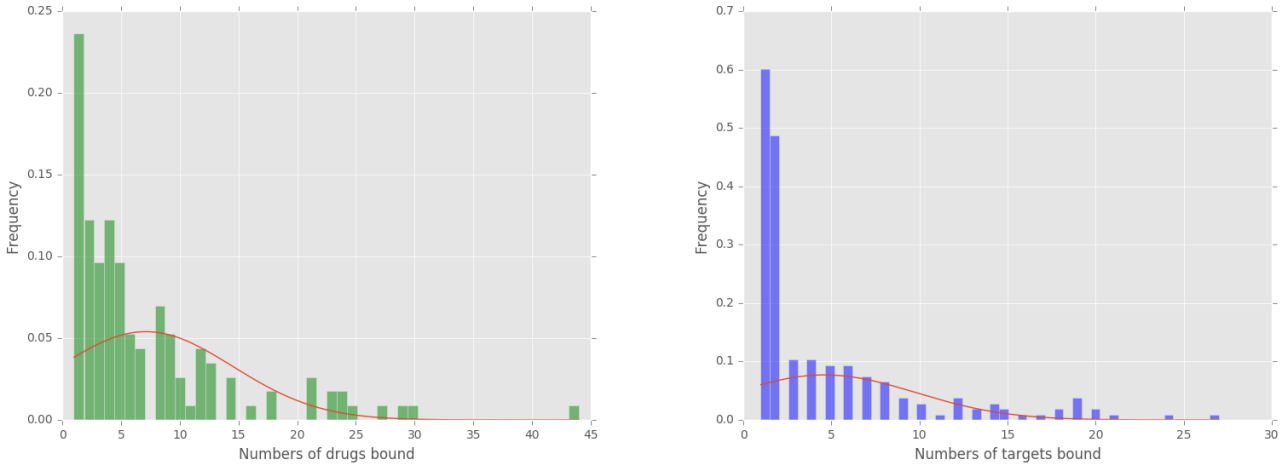
We observed no statistically significant difference between the DS-Score of the FDA-targets and the GSK-PKIS targets (pvalue = 0.3699). This implies that targets of both set tends to form similar network communities. However, as reported by several studies, we also found that the tyrosine kinase family are the most targeted family by the the approved drug with a coverage of over 70% depicted by our drug-target dataset. Since the GSK-PKIS had more compounds compared to the FDA-drugs, we therefore chose this as our "test

set" to analyse the promiscuity of protein kinase inhibitors and possible repurposing of the drugs by associating the targets to the domain families (Pfam-FunFams).

[We choose the FDA set as a reference set not because it has the same tendency to form communities in the network as the GSK set, but because the FDA set comprises effective, marketed drugs that target protein kinases with proven therapeutically applications. This result should be just an indication that the GSK set contain compounds that might be as good drugs as the real known anti-protein kinase drugs]. [The distinction between the reference set and the testing set should me made clear before this, when you present both datasets. In fact, we should elevate the comparison of the two sets as a main aspect of the analysis, through the comparison of the network properties of both sets we ca suggest that the GSK set could make good anti-kinase drugs] [Also, you should use the matrix similarity too, to compare both sets, not only the ds score] [Also, we should add the other network promerty in the comparison between the two sets: betweenness centrality]

Target Promiscuity of Protein Kinase Inhibitors

The PKIS dataset obtained from ChEMBL database was analysed to study the association of kinases (targets) with inhibitors in order to understand the inherent promiscuity associated with the protein kinase inhibitors set.



(a) Promiscuity of drugs for kinases

(b) Promiscuity of targets for PKIS

Figure 4: Distribution of kinase-drug interaction and drug-kinase interaction for the GSK-PKIS sets

Figure 4 shows the inherent promiscuity of the kinases for inhibitors as well as inhibitors for kinases. Protein kinase inhibitors interact with more than one kinase while some kinases could also interact with more than one protein kinase inhibitor, which indicates the non-specificity of these kinase sets. This is not surprising as the majority of the protein kinase inhibitors are the ~~clasical~~ classical type 1 inhibitors that binds at the ATP site directly and compete with ATP which is quite universal and and what?

[Are the FDA drugs promiscuous too? Compare the promiscuity of the FDA drugs with the GSK compounds and discuss similarities differences]

[Before adding the FunFams to the mix we should establish the promiscuity and network properties of the GSK compounds (i.e. network aggregation and betweenness centrality). We do this by comparing the GSK compounds with the FDA drugs. Then we can discuss FunFams etc.] [Also, since we need the CATH-FunFams ready to perform the bulk of the work, you can use the time in establishing these comparisons (FDA-GSKA) and discussing the network properties of the GSK compounds and GSK drugs.]

Mapping kinases to Pfam-FunFam

Protein kinases have been divided divided into 9 groups by [Manning et al., 2002] as explained in the introductory section. We associated all human kinases with our functional families (groups of evolutionarily related, structurally and functionally coherent protein families). We scanned all human kinase sequences from Pfam against the in-house Pfam-FunFams library using HMMer3. Pfam-FunFam data from the Gene3D database was used as Pfam provides sequences that cover the entire kinase catalytic region, whilst CATH divides the kinases into the N and C lobe domains.

The identified 1277 human protein kinases sequences obtained from Pfam were clustered at 90% sequence identity to remove protein isoforms and obtain a unique non-redundant human kinase set using CD-HIT clustering algorithm [Li and Godzik, 2006]. The 741 non-redundant kinases obtained were distributed amongst 130 Pfam-FunFams which map to 16 of the 35 Pfam clans. We mapped the Pfam-FunFam to the human kinome tree and found that our domain-family classification correspond well to the various groups with no overlaps between the groups identified by [Manning et al., 2002] as shown in figure 5

below.

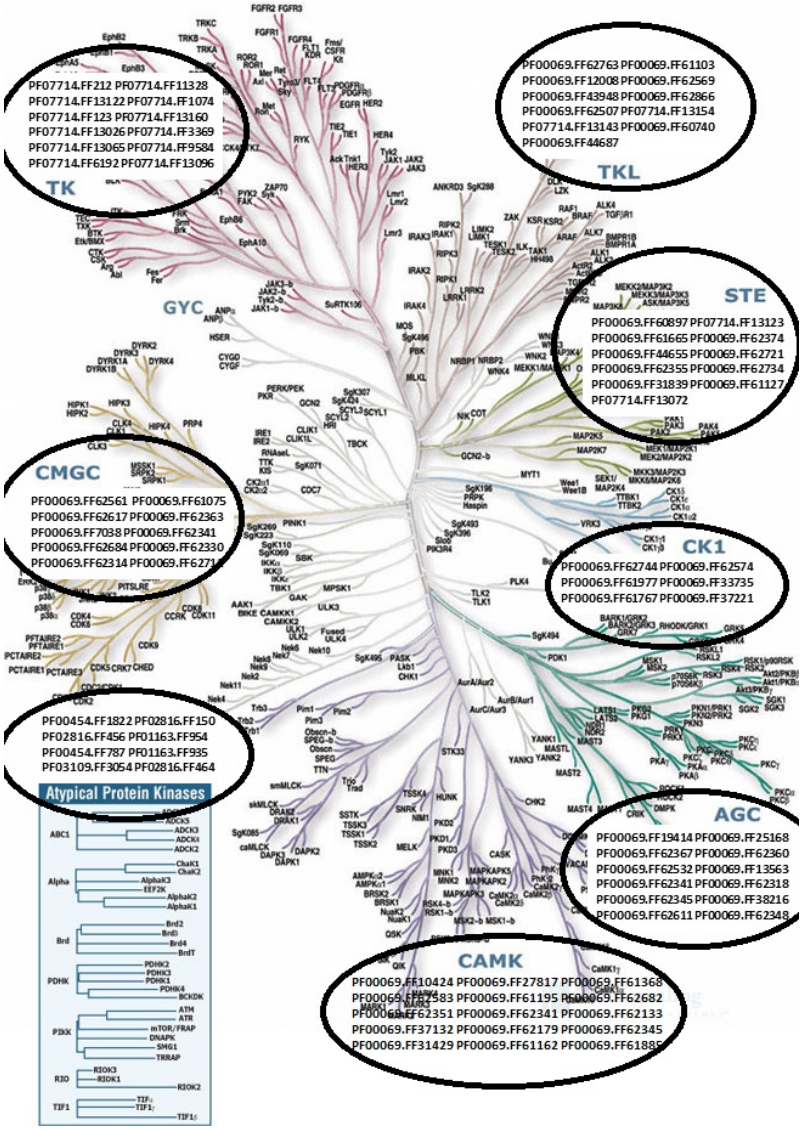


Figure 5: Pfam-FunFam distribution across the human kinome.The kinome tree was adapted from [Manning et al., 2002]. The mapping of Pfam-FunFam to the group "Others" is not shown in this figure

We then mapped the drug-targets of the PKIS obtained from ChEMBL to the Pfam-FunFams at 50% inhibition level with an affinity of 0.1μM and found that they are associated with 37 of the 130 Pfam-FunFams which represents about 30% of the human kinase Pfam-FunFams obtained. We speculated that this 30% represents the kinases most targeted by the pharmaceutical industry based on their relevance to human disease and those most studied as the current research in kinase therapeutics indicates that only 10-15% of the kinases are being targeted [Li et al., 2016, Elkins et al., 2016]. We also observed this ratio

from our kinase-inhibitor set being used in this study.

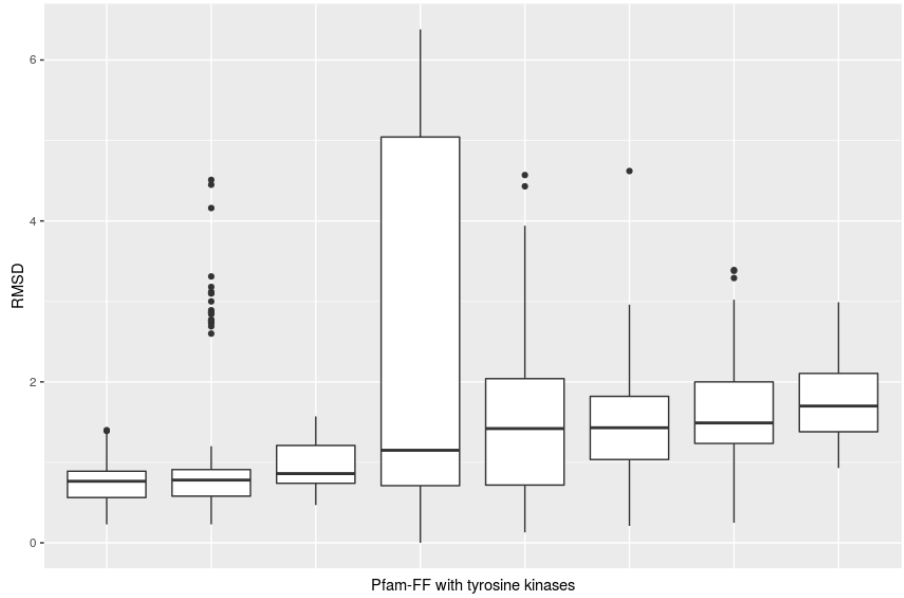
Table 1: Pfam-FunFam families shared by the GSK and FDA dataset

Pfam-FunFams	GSK	FDA	GSK-targets	FDA-targets	Shared-targets
PF00069.FF62355	1	2	2	16	2
PF00069.FF62314	4	1	3	6	2
PF00069.FF62345	5	1	7	12	5
PF07714.FF212	9	1	3	3	3
PF07714.FF13154	2	2	1	3	1
PF07714.FF13026	17	4	25	36	25
PF07714.FF13065	4	5	8	17	8
PF07714.FF123	23	7	5	5	5

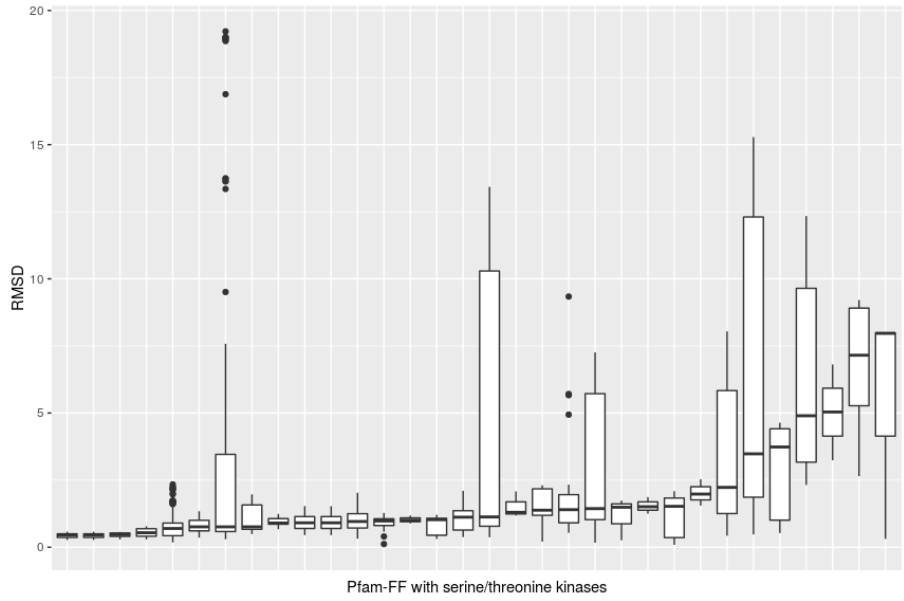
Table 1 shows the list of the targets shared by the Pfam-FunFams that are shared by the GSK and FDA approved drugs. Although, the FDA approved drugs are quite small compared to the GSK-PKIS, we however observed a higher number of targets associated with them. This creates room for possible repurposing of the experimental drugs (GSK) to other targtes within the family.

Structural conservation of human kinase relatives in the Pfam-FunFams

We measured how structurally conserved the kinases are across a given Pfam-FunFam. The human relatives of the Pfam-FunFam were mapped to structure by using the SIFT mapping of UniProt sequences to PDB, while the domain regions were specified using Pfam. The structures were then evaluated for structural conservation by measuring RMSD following the pairwise alignment by the SSAP algorithm [Orengo and Taylor, 1996] and the superposition by ProFit [Martin, 2009]. The groups were divided into the tyrosine-kinase Pfam-FunFams and serine/threonine Pfam-FunFams. The distribution of the RMSD across the two Pfam-FunFam groups is shown in figure 6 below.



(a) RMSD across the Tyr Pfam-FunFam



(b) RMSD across the Ser/Thr Pfam-FunFam

Figure 6: Structural coherence and conservation of the Pfam-FunFam kinases measured using the SSAP-algorithm

Figure 6 suggest significant structural conservation of relatives in these families. As for many comparisons, the observed RMSD was below 2 in both the tyrosine kinases as well as the serine/threonine kinase FunFams. This supports the overall view of structural conservation of the kinases in the literature [Manning et al., 2002, Taylor and Kornev, 2011, Elkins et al., 2016, Roskoski, 2016]. However, in some Pfam-FunFams pairs of relatives

have a high RMSD score which tends to distort the overall RMSD measure, indicating lack of structural coherence of relatives in these FunFams.

A deeper insight into one of such families was carried out using the Pfam-FunFam PF07714.FF13122 which gave an overall RMSD below 2A but contains outliers having RMSD as high as 5A. The reason for this could be attributed to the multidomain architecture of the relatives of the Pfam-FunFam as majority of the relative in this protein have additional SH2-domain while others either have an immunoglobulin-domains or no additional domain.

Enrichment test of Pfam-FunFams associated with drug targets

The Published Kinase Inhibitor Sets (PKIS) was identified as the chemical starting point for probing orphan kinases. The study by [Anastassiadis et al., 2011] illustrated the utility of these compounds for developing selective inhibitors against untargeted kinases LOK and SLK . Thus, the use of domain families could help increase the coverage of potential targets of the kinase-inhibitor set as the targeted kinases are about 10-15% . We used our FunFam-target enrichment analysis protocol (see chapter 2) to test for the overrepresentation of drugs in our FunFams.

This potocol uses a binomial test followed by multiple testing for correction to determine the most appropriate FunFam with which a Kinase inhibitor associates. We found that 109 PKIS-drugs were overrepresented in 30 Pfam-FunFam at p-value ≤ 0.05.Figure 7 shows the distribution of these 30 FunFams and the numbers of drugs they are associated with.

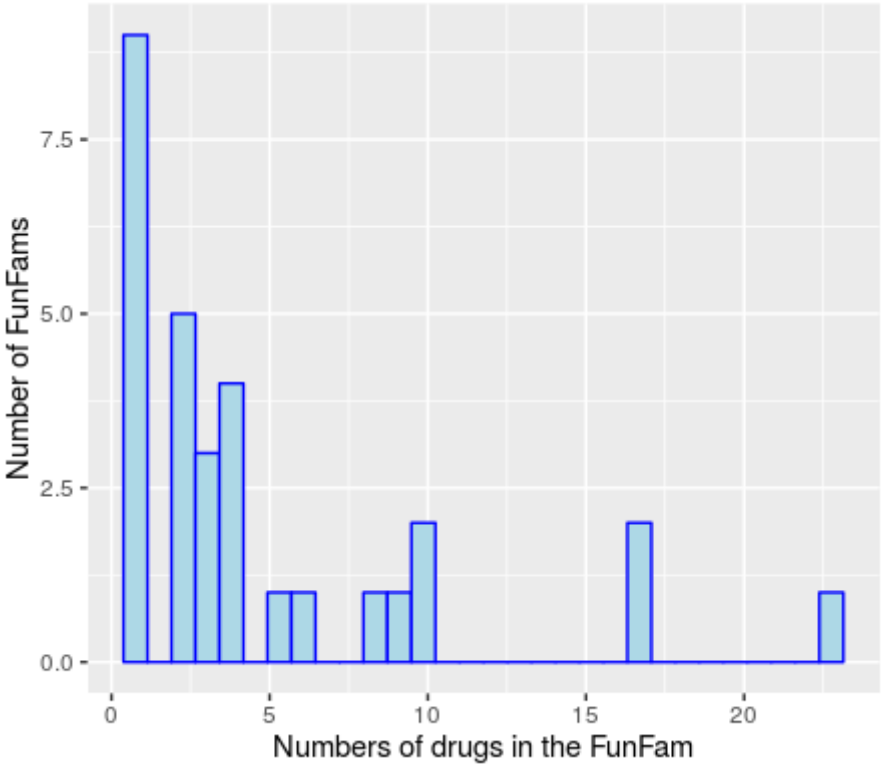


Figure 7: Distribution of drugs associated with FunFams

As shown in figure 7, we observed that 70% of the enriched Pfam-FunFams were associated with more than 2 kinase inhibitors from the PKIS set. This enriched Pfam-FunFam dataset was used for further analysis of the protein-kinase inhibitor targets on protein-protein interaction network.

The Pfam-FunFam-drug interaction was represented in network for visualization in Cytoscape as shown in figure 8 while the names of the overrepresented Pfam-FunFams were identify using our in-house program for the UniProt description of protein sequences as shown in table **??**.

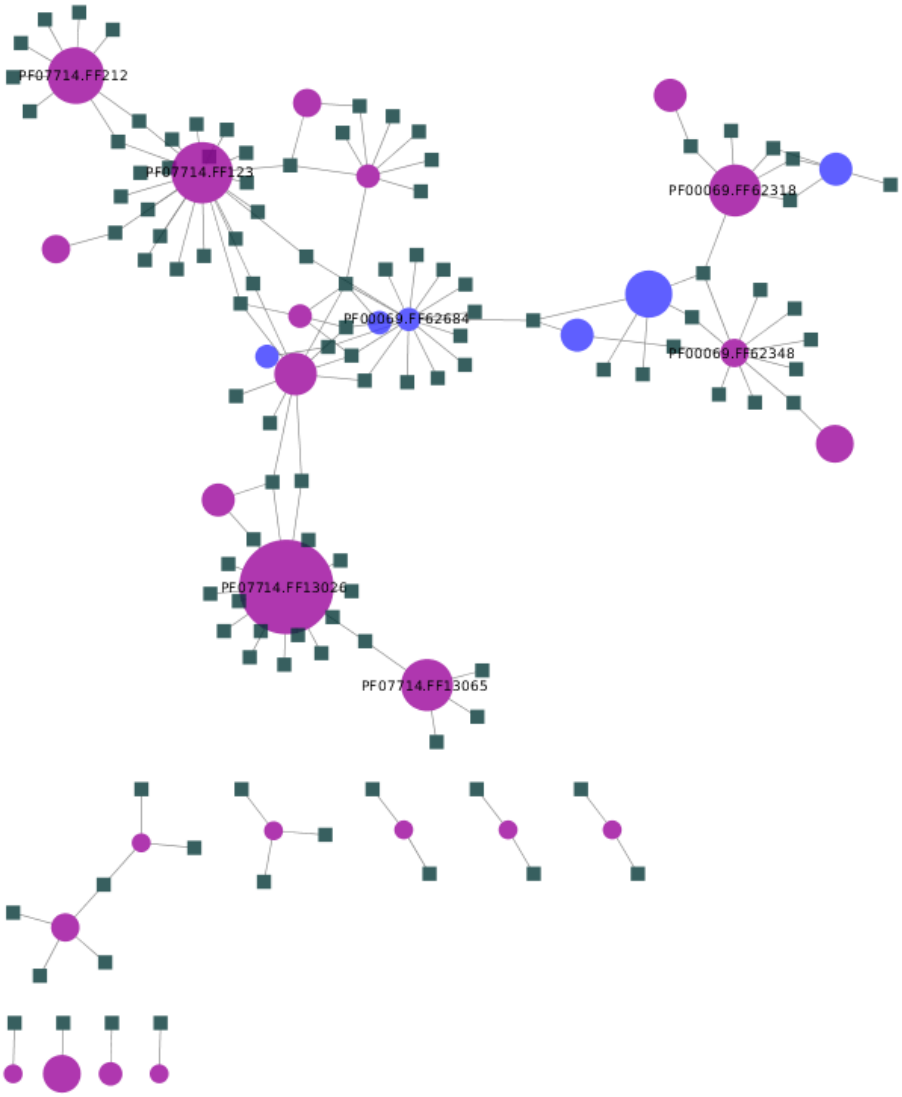


Figure 8: Pfam-FumFam-drug interaction network. In this network, the green coloured square node represents the CheMBL protein kinase inhibitors while the purple coloured circle nodes are the Pfam-FunFam whose relative are structurally coherent with a mean RMSD$\leq$ 2 while the blue coloured circle are Pfam-FunFam with RMSD value $\geq$ 2.5. The size of each node reflects the numbers of targets (relatives) in each family. Also labelled are some families with relatives $\geq$ 5 and interacting with at least 5 drugs.

Network Analysis of the Pfam-FunFams Enriched with Drug Targets

The representation of proteins on a network gives a view of the information flow and interactors for biological process. We obtained a functional protein association network for human from the STRING database version 10.0 [Szklarczyk et al., 2014]. We chose

STRING database as it is widely used and oftentimes updated. STRING database generate scores to measure reliabilty of an interaction by benchmarking predicted interactions against common set of true positive associations. We filtered this data by applying a cut-off of 800 on the combined score of the interaction which correspond to those PPI with high reliabilty. This gave 219,608 physical interactions between 10,430 proteins. We extracted the largest connected subgraph and then computed the node centralities for drug targets as well as the enriched Pfam-FunFams using centrality measures such as the betweenness centrality.

Centrality Measure

Centrality measures identify important nodes relative to other nodes within the network. Such measures include the degree, closeness centrality, betweenness centrality as well as the PageRank. The degree of a node is the number of connections (edges) it shares with other nodes. The betweenness centrality (BC) is the fraction of the number of shortest paths that pass through each node. The BC measures how often a node occurs on all the shortest paths between two nodes. Therefore, a node with high BC influences the flow of information in the network.

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$$

where V is the set of nodes, $\sigma(s,t)$ *is the number of shortest (s, t)-paths, and* $\sigma(s,t|v)$ *is the number of those paths passing through some node v other than s, t. If s = t,* $\sigma(s,t) = 1$*, and if* $v \in s,t$*,* $\sigma(s,t|v) = 0$

Centrality measures and similarities of kinase inhibitor targets

We measured the topological properties of the targeted kinases and compared them to other human kinases (excluding the targeted kinases) and all proteins in a given human functional network. we defined here ”Hubs” as the top 20% of nodes with the highest degree (connections) while ”High-BC” are those top 20% of nodes with the highest betweenness centrality. The ”Bottlenecks” are defined as those nodes within the ”High-BC” but excluded from ”Hubs” i.e. those with low degree connectivity. The ”Hubs & High-BC” are those set of nodes in the Hubs group with high betweenness centrality measure.

The measures of degree and betweenness centrality are amongst the most profound topological properties of nodes in a protein interaction network. The bottlenecks in a protein functional network represents key connectors and show functional and dynamic properties [Yu et al., 2007]. Table 2 shows the comparison of the targeted kinases, all human kinases and all human proteins in a functional network. The proportion is a measured relative to each group as;

$$\frac{\text{Number of Hubs or bottlenecks in targeted set} \times 100}{\text{Total number proteins in targeted set}}$$

Table 2: Topological analysis of kinases in a functional protein network

Groups	*Hubs (%)*	*Bottlenecks (%)*	*High-BC (%)*	*Hubs & High-BC (%)*
Targeted kinases	*42.06*	*5.61*	*41.12*	*46.73*
Other-human kinases	*19.59*	*6.76*	*16.44*	*23.20*
All-human proteins	*19.19*	*11.51*	*7.47*	*18.98*

The results therefore show that only a small percentage of the targeted kinases are bottlenecks. The kinases are most likely to be hubs and they are highly central in protein functional network with high betweenness centrality (Mannwhitney test for hubs and centrality between kinases and other proteins in a network; pvalue = 2.54e-13 amd 6.617e-25). We hypothesize that bottlenecks will be a good drug target as they are central in the network but associated with less nodes (thus less functional disintegration expected). However, there is still a lot of debate on this opinion as some studies have suggested bottlenecks to be associated with side effects. For instance, in the studies by [Perez-Lopez et al., 2015], they observed that targets of drugs with side effects are better spreader of pertubation in the interactome indicating that these targets are quite central and they disrupts the interactome more as compared to the drug-targets without side effects or non-target proteins.

Dispersion measure of targets of kinase inhibitor in a protein network

We also investigated the behaviour of kinase inhibitor targets in human protein interaction network and measured the dispersion of the targets in network using dispersion measures such as the matrix similarity as well as the shortest path distance across all the targets of a

drug a measure which was adapted from the study conducted by [Menche et al., 2015] to study disease genes in network.

We selected all the targeted proteins from ChEMBL 23 and excluded the kinases whilst we grouped all kinase drugs together (i.e. GSK-PKIS and the FDA approved drug). We also consider less specific interaction between drugs and proteins as a way of classifying off-targets using a pchembl<6.

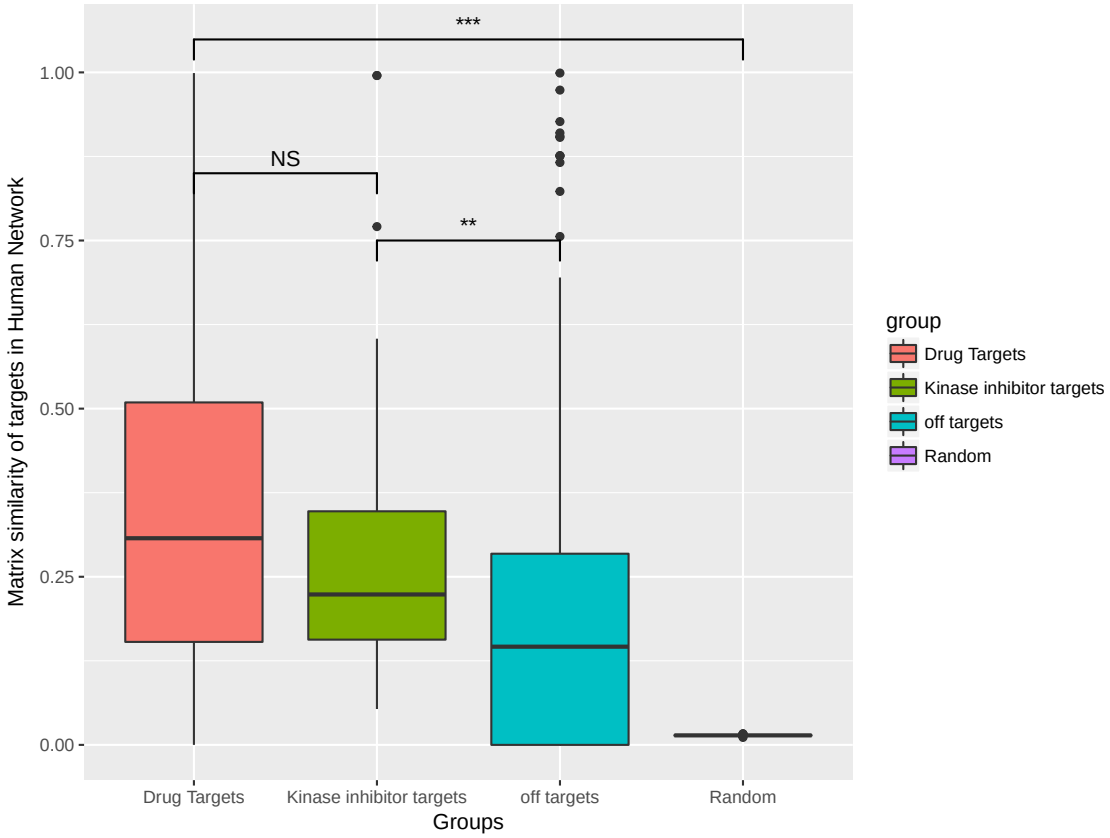


Figure 9: Box plot comparing the distribution of the matrix similarity of the various group against random. (NS indicate not statistically significant, while ** indicate statistical significance between pairs)

The result in figure 9 show that there is no difference in the observed matrix similarity of the kinase-targets and other drug targets (with pvalue = 0.9113). This observation is consistent with our previous report about drug targets having higher similarities in the protein functional network [Moya-García et al., 2017]. However when we compared the matrix similarities of the drug targets as well as the kinase inhibitor targets with drug off-targets, we found a statistically significant difference of the observed matrix similarities (pvalue = 2.09e-10 and 4.419e-9) for the comparison of drug off-target similarity with

drug-targets and kinase inhibitor targets respectively.

Using the [Menche et al., 2015] score called "DS-Score" which measures the mean distance of seperation of genes within a cluster. It is expected that the lower the "DS-Score", the more aggregated the proteins are in the network. We also used this measure to compare the drug targets aggregation in a given functional network.

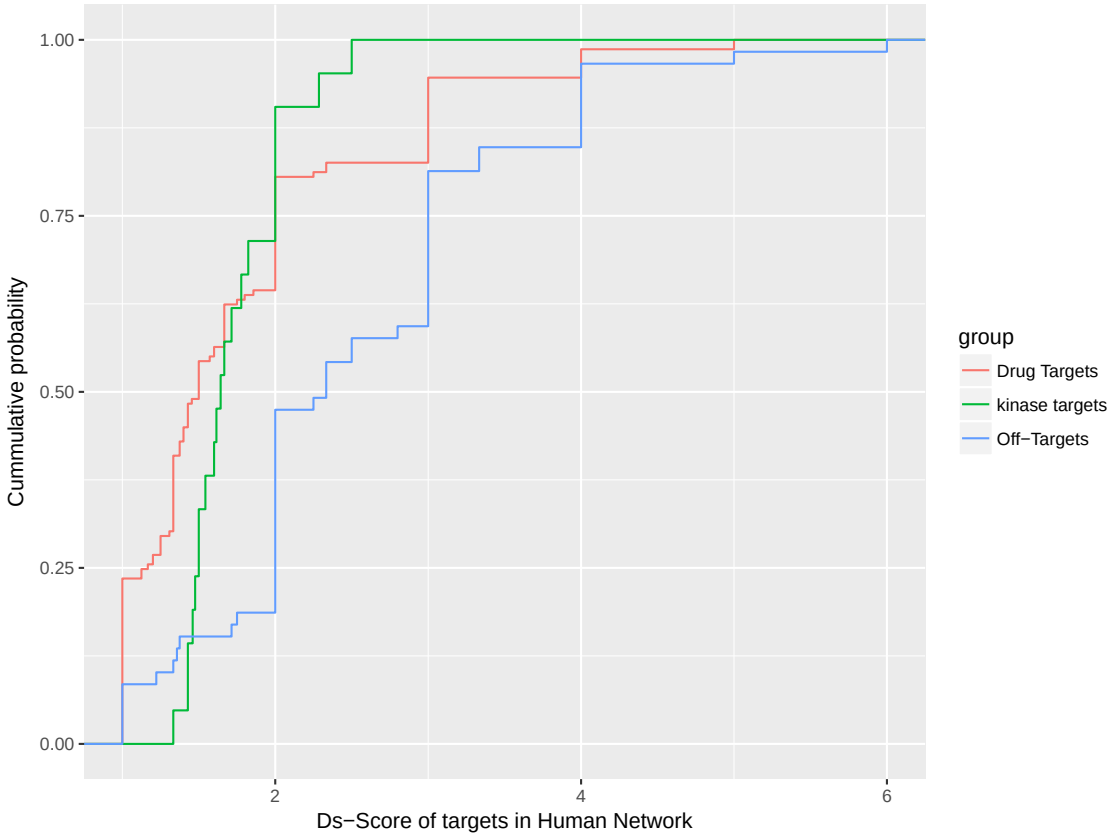


Figure 10: The cummulative probability plot comparing "DS-Score" of the drug targets, kinase inhibitor targets and off-targets.

The result (figure 10), show that the DS-Score is higher in offtargets as compared to the drug targets as well as the kinase targets and this variation is statistically significant (pvalue = 7.035e-8 and 0.0001542. There is however no statistically significant difference between the DS-S35core of the drug targets and kinase inhibitor targets (p=0.1447). This result therefore show that the offtargets are dispersed in the network as compared to the targeted proteins which are aggregated in the network. We have therefore shown using two different measures "matrix similarity" and "DS-Scores" that targets are clustered in the network as compared to offtargets or random proteins. Previous studies

have also considered diseased proteins to be clustered in modules in protein network [Menche et al., 2015], if we therefore considered the disease proteins as the drug targets in a network, the aggregation which we observed in our study therefore implies that drug targets behave like disease protein in the human protein network.

Dispersion of Drug Targets in Pfam-FunFams in a Protein Functional Network

Following our initial hypothesis that FunFams whose relatives aggregate in a network neighbourhood are likely to be enriched with potential targets and free of off-targets, we assessed the matrix similarity of all the Pfam-FunFam kinases. The matrix similarities have a score ranging from 0-1 with 1 indicating a high similarity, i.e. close proximity in the network and 0 showing no similarity.

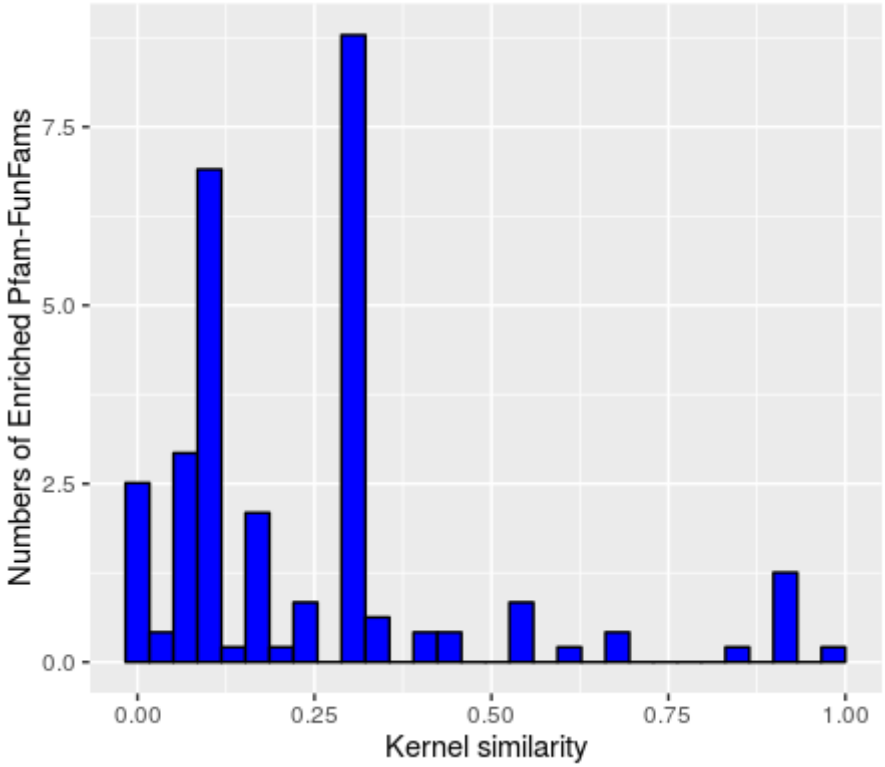
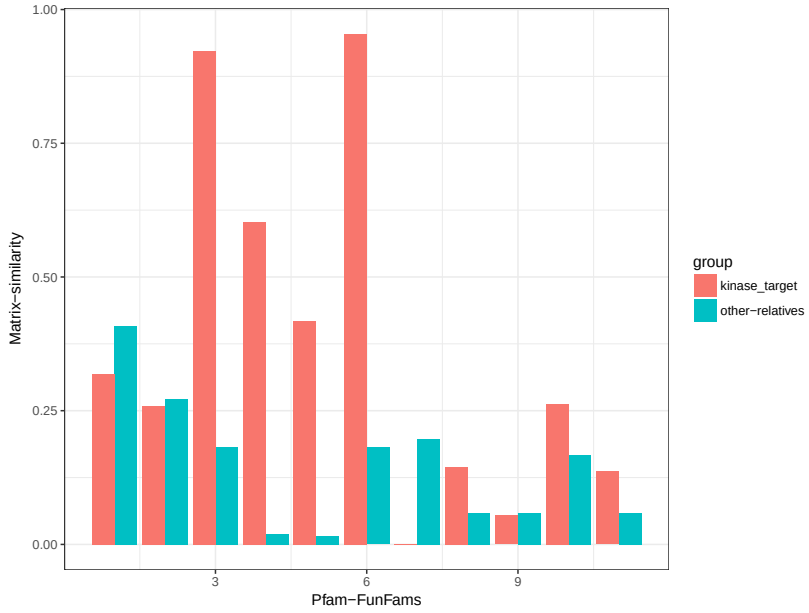


Figure 11: Distribution of the matrix similarity measure across the Pfam-FunFam.

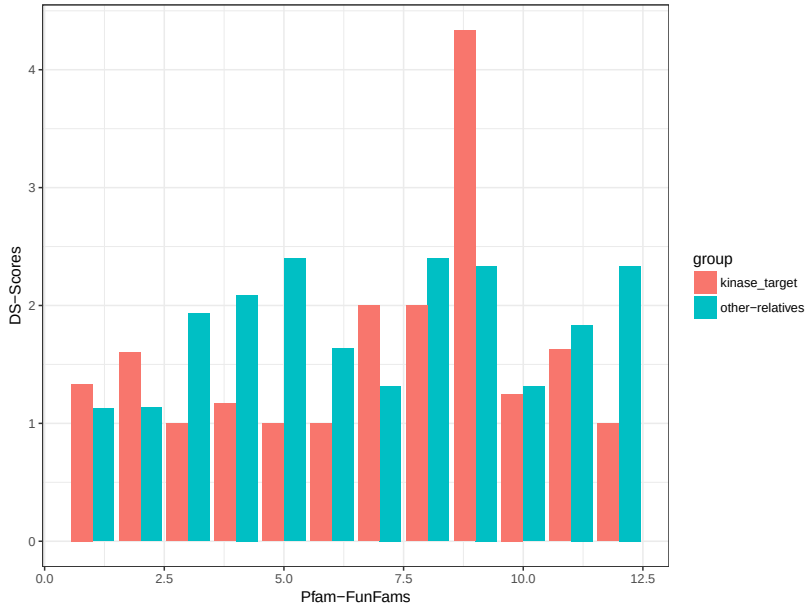
The results indicate that most of the Pfam-FunFam relatives are scattered across the functional protein network with majority of the Pfam-FunFam family having a similarity score lower than 0.5. Although we've have not shown a given threshold for the cutoff for prediction of side effect, but majority of the kinases have a matrix similarity lower than

0.5 and this may explain the side effects associated with protein kinase inhibitors as their targets are quite distributed in the network and could affect many biological pathways i.e. elicits multiple pharmacological responses in a given organism.

We then compared the matrix similarity of targeted kinase in each FunFam against other relatives of the same family to compare the network properties of targets and other non-targeted relatives of the same FunFam.



(a) Bar-plot of the matrix similarity of targets in each family compared to other non-targeted relatives



(b) Bar-plot of the DS-Scores of targets in each family compared to other non-targeted relatives

Figure 12: Similarity measure across the Pfam-kinase family measured using the matrix similarity and DS-Scores

As shown in figure 12 the targeted kinases have a higher matrix similarity compared to other relatives of the same FunFam. The same observation was also found using a different measure (DS-Score) which indicate that majority of the kinase target are close in

the network with lower DS-Score as we have earlier shown that the lower the DS-Score, the more aggregated a set of proteins are in the network [Menche et al., 2015]

Structural coherence of the binding site of the enriched Pfam-FunFam relatives

The work published by [Elkins et al., 2016] reported some solved structures for the binding of inhibitors from the PKIS with some kinases. Crystal structures of the inhibitor (CHEMBL237571) bound to the lymphocyte-oriented kinase (LOK) in the inactive DFG-out state (PDB-ID:4USD) and active DFG-in state (PDB-ID: 4USE) have been deposited in PDB. Using this example, we examined the binding site conservation across members of the Pfam-FunFam this target belongs to.

The target for this inhibitor belongs to the Pfam-FunFam PF00069.62355 (a STE-group kinase) that has about 80 relatives. Using the SIFTS mapping of UniProt-sequences to PDB structures, we were able to identify 15 members (about 18%) of this family with PDB structures. In case of multiple structures of the same kinase, firstly, we find an unbound structure, or if there are no unbound structures then we chose the best resolved structure for the particular kinase.

Table 3: The list of Pfam-FunFam (PF00069.62355) relatives with structural information

UniProt-ID	PDB-ID	Resolution (Å)
O95819	4u3y	1.45
Q9P289	3ggf	2.35
Q9P286	2f57	1.80
O96013	2j0i	1.60
Q8IVH8	5j5t	2.85
Q9Y6E0	3a7f	1.55
Q13153	1yhv	1.80
Q9UKE5	2x7f	2.80
Q9H2G2	2j51	2.10
Q9NQU5	2c30	1.60
Q99759	2o2v	1.83
Q9Y2U5	5ex0	2.70
Q13177	3pcs	2.86
O00506	2xik	1.97
O94804	2j7t	2.00

We structurally superposed all the members of this family. This family was found to be structurally coherent as we used ProFit algorithm guided by SSAP alignnment and obtained an average RMSD of 1.11 ± 0.48.

Figure 13: Structural alignment and superposition of the relatives in Pfam-FunFam (PF00069.62355) based on the alignment of the binding region. The interacting residues are coloured in yellow while the secondary structures are coloured accordingly (Beta sheet(blue), alpha (Magenta), the inhibitor is coloured in rainbow.

Network analysis of MutFams enriched with kinases

We considered the network properties of mutationally enriched FunFams (MutFams). These families are enriched in cancer mutations (taken from the COSMIC database [Bamford et al., 2004]. We compared these network properties with FunFams enriched in non-disease associated neutral variations(mutation data taken from HumVar [Capriotti et al., 2006]). This approach was taken to find out whether the kinases linked to diseases are more dispersed in the netweork making it hard to target them with drugs.

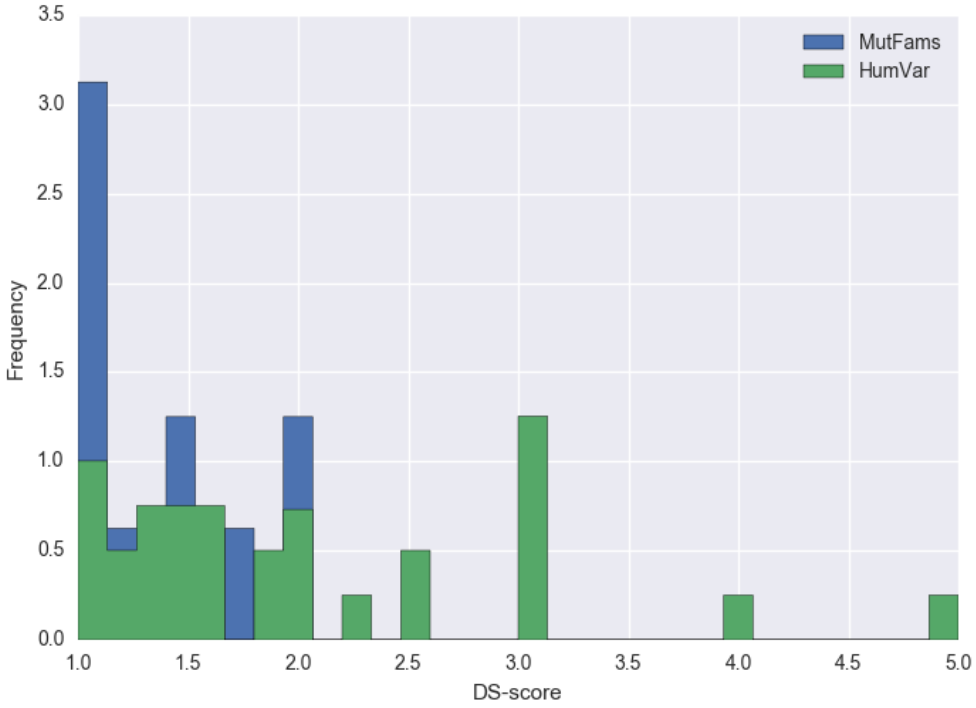


Figure 14: DS-measure of the MutFam in comparison with the HUMVAR

Figure 14 shows a plot of the distribution of the DS-score between MutFams and HumVar. There is a significant difference between these two sets of genes as the relatives of MutFams are highly clustered in the human functional network compared to the relatives in FunFams ($P = 0.00645$). MutFams therefore provides a reasonable annotation of disease genes with lower side effects which makes them suited to targetting by inhibitors.

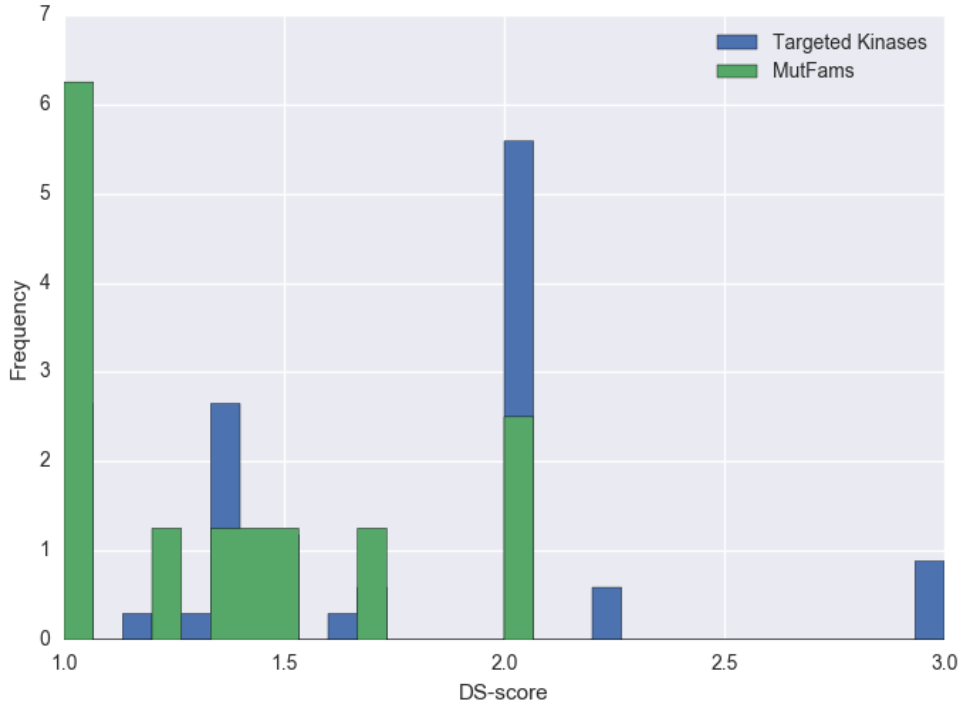


Figure 15: DS-measure of the MutFam in comparison with the targeted kinases

The DS-score of the MutFams were compared with the targeted kinases (figure 15), these different sets of genes tend to be clustered in the same fashion as there is no difference in the observed DS-score of the MutFams and the targeted kinases ($P = 0.0191$). This indicate that the targeted kinases share similar network characteristics as the mutated sets of genes that are implicated in human diseases. These MutFams are therefore potential therapeutic targets that could be harnessed and considered for therapeutic purposes.

Table 4: The mutfam classes and their representation in Pfam-FunFams with the similarity measure in the protein functional network

Cancer Types	CATH-FunFam	Pfam-FunFam	%overlap	No of Drugs	matrix-sim	DS-score
LGG	1.10.510.10.FF78531	PF07714.FF13154	47.1	2	0.846	1.2
LGG	1.10.510.10.FF79008	PF00069.FF27817	46.2		0.541267	1
BLCA	1.25.40.70.FF2223	PF00454.FF1812	43.2		0.662619	1.5
BRCA	3.30.200.20.FF2866					
BRCA	2.30.29.30.FF22238	PF00069.FF25168	14.2	4	0.319	1
BRCA	1.10.1070.11.FF1687	PF00454.FF1812	37.3		0.662619	1.5
BRCA	1.25.40.70.FF2223	PF00454.FF1812	43.2		0.662619	1.5
COAD	3.30.200.20.FF64824					
COAD	1.20.120.330.FF23932					
COAD	1.10.510.10.FF78531	PF07714.FF13154	47.1	2	0.846	1.2
COAD	1.10.1070.11.FF1687	PF00454.FF1812	37.3		0.662619	1.5
COAD	1.10.510.10.FF79298	PF00069.FF62569	100		0.584709	1
COAD	1.25.40.70.FF2223	PF00454.FF1812	43.2		0.662619	1.5
COAD	1.10.510.10.FF78966	PF07714.FF3369	35.7	1	0.097	
COAD	1.10.510.10.FF79140	PF07714.FF13122	100		0.220333	1
GBM	1.10.510.10.FF79478	PF00069.FF61939	45.5		0.694	1
GBM	3.30.505.10.FF4305				0.509667	
GBM	1.10.510.10.FF79008	PF00069.FF27817	46.2		0.541267	1
GLI	1.10.510.10.FF78531	PF07714.FF13154	47.1	2	0.846	1.2
GLI	1.10.510.10.FF79478	PF00069.FF61939	45.5		0.694	1
GLI	1.10.510.10.FF79008	PF00069.FF27817	46.2		0.509667	1
GLI	3.30.505.10.FF4305					
GLI	1.25.40.70.FF2223	PF00454.FF1812	43.2		0.662619	1.5
KIRC	1.25.40.70.FF2223	PF00454.FF1812	43.2		0.662619	1.5
LAML	1.10.510.10.FF78745	PF07714.FF13026	53.1	17	0.184	1.2
LIHC	1.25.40.70.FF2223	PF00454.FF1812	43.2		0.662619	1.5
LIHC	3.30.60.20.FF5564	PF00069.FF62318	23.9	6	0.3411	1.17
LUAD	3.30.200.20.FF1240					
LUAD	3.30.200.20.FF64824					
LUAD	1.10.510.10.FF79008	PF00069.FF27817	46.2		0.541267	1
LUAD	1.10.510.10.FF79228	PF00069.FF62599	71.4		0.588333	1.5
LUSC	1.25.40.70.FF2223	PF00454.FF1812	43.2		0.662619	1.5
PAAD	1.10.510.10.FF78763	PF00069.FF62351	43.5	1	0.155692	1.5
READ	1.10.510.10.FF78531	PF07714.FF13154	47.1		0.846	1.2
READ	1.10.1070.11.FF1687	PF00454.FF1812	37.3		0.662619	1.5
READ	1.10.510.10.FF78946	PF00069.FF62345	43.8	5	0.4562	1.72
SKCM	1.10.510.10.FF78531	PF07714.FF13154	47.1	2	0.846	1.2
THCA	1.10.510.10.FF78531	PF07714.FF13154	47.1	2	0.846	1.2
UCEC	1.25.40.70.FF2223	PF00454.FF1812	43.2		0.662619	1.5

References

[Anastassiadis et al., 2011] Anastassiadis, T., Deacon, S. W., Devarajan, K., Ma, H., and Peterson, J. R. (2011). Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. Nature biotechnology, 29(11):1039–1045.

[Bamford et al., 2004] Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P. A., Stratton, M. R., et al. (2004). The cosmic (catalogue of somatic mutations in cancer) database and website. British journal of cancer, 91(2):355.

[Capriotti et al., 2006] Capriotti, E., Calabrese, R., and Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics, 22(22):2729–2734.

[Dranchak et al., 2013] Dranchak, P., MacArthur, R., Guha, R., Zuercher, W. J., Drewry, D. H., Auld, D. S., and Inglese, J. (2013). Profile of the gsk published protein kinase inhibitor set across atp-dependent and-independent luciferases: implications for reporter-gene assays. PloS one, 8(3):e57888.

[Elkins et al., 2016] Elkins, J. M., Fedele, V., Szklarz, M., Azeez, K. R. A., Salah, E., Mikolajczyk, J., Romanov, S., Sepetov, N., Huang, X.-P., Roth, B. L., et al. (2016). Comprehensive characterization of the published kinase inhibitor set. Nature biotechnology, 34(1):95.

[Gaulton et al., 2016] Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrián-Uhalte, E., et al. (2016). The chembl database in 2017. Nucleic acids research, 45(D1):D945–D954.

[Knapp et al., 2013] Knapp, S., Arruda, P., Blagg, J., Burley, S., Drewry, D. H., Edwards, A., Fabbro, D., Gillespie, P., Gray, N. S., Kuster, B., et al. (2013). A public-private partnership to unlock the untargeted kinome. Nature chemical biology, 9(1):3–6.

[Li and Godzik, 2006] Li, W. and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 22(13):1658–1659.

[Li et al., 2016] Li, Y. H., Wang, P. P., Li, X. X., Yu, C. Y., Yang, H., Zhou, J., Xue, W. W., Tan, J., and Zhu, F. (2016). The human kinome targeted by fda approved multi-target

drugs and combination products: a comparative study from the drug-target interaction network perspective. PloS one, 11(11):e0165737.

[Manning et al., 2002] Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science, 298(5600):1912–1934.

[Martin, 2009] Martin, A. (2009). Profit program using mclachlan algorithm. http://www.bioinf.org.uk/software/profit.

[Menche et al., 2015] Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., and Barabási, A.-L. (2015). Uncovering disease-disease relationships through the incomplete interactome. Science, 347(6224):1257601.

[Moya-García et al., 2017] Moya-García, A., Adeyelu, T., Kruger, F. A., Dawson, N. L., Lees, J. G., Overington, J. P., Orengo, C., and Ranea, J. A. (2017). Structural and functional view of polypharmacology. Scientific Reports, 7.

[Orengo and Taylor, 1996] Orengo, C. A. and Taylor, W. R. (1996). [36] ssap: sequential structure alignment program for protein structure comparison. Methods in enzymology, 266:617–635.

[Perez-Lopez et al., 2015] Perez-Lopez, Á. R., Szalay, K. Z., Türei, D., Módos, D., Lenti, K., Korcsmáros, T., and Csermely, P. (2015). Targets of drugs are generally, and targets of drugs having side effects are specifically good spreaders of human interactome perturbations. Scientific reports, 5:10182.

[Roskoski, 2016] Roskoski, R. (2016). Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. Pharmacological research, 103:26–48.

[Szklarczyk et al., 2014] Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., et al. (2014). String v10: protein–protein interaction networks, integrated over the tree of life. Nucleic acids research, 43(D1):D447–D452.

[Taylor and Kornev, 2011] Taylor, S. S. and Kornev, A. P. (2011). Protein kinases: evolution of dynamic regulatory proteins. Trends in biochemical sciences, 36(2):65–77.

[Yu et al., 2007] Yu, H., Kim, P. M., Sprecher, E., Trifonov, V., and Gerstein, M. (2007). The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS computational biology, 3(4):e59.