**Methods**:

From the results of the random search optimization, we conducted a series of one-way ANOVAs to determine whether database selection had a significant impact on each step of X2K (TFEA, PPI-construction, Kinase-enrichment). In each ANOVA, the dependent variable was which databases were selected, while the independent variable was the score, defined as -log(P-value) of the target kinase. Two additional one-way ANOVAs were conducted to test the effect of ‘number of PPI subnetwork nodes’ on the score, as well as ‘number of significant kinases’ on the score.

**Results**:

The selection of TF databases did have a significant impact on the score (ANOVA: P-value < 0.0001) (Table 1). ‘ChEA & ENCODE Consensus’ was associated with the highest average score (when the target kinase was correctly predicted) (Figure 1), however ‘Transfac & Jaspar’ was associated with most often correctly predicting the target kinase, as opposed to missing it entirely (23.38% of the time) (Figure 6).

The selection of PPI databases did have a significant impact on the score (ANOVA: P-value < 0.0001, degrees of freedom) (Table 1).

‘Stelzl’ was associated with the highest average score (when the target kinase was correctly predicted) (Figure 2), however ‘PPID’ was associated with most often correctly predicting the target kinase (33.80% of the time) (Figure 7). PPID, IntAct, MINT and BioGRID followed PPID in terms of the highest average scores.

The selection of kinase databases did not have a significant impact on the score (ANOVA: P-value ≥ 0.05) (Table 1). ‘KEA 2018’ was associated with the highest average score, (when the target kinase was correctly predicted) (Figure 3), but ‘ARCHS4’ was associated with most often correctly predicting the target kinase (17.95% of the time) (Figure 8).

The ‘number of PPI subnetwork nodes’ did not have a significant effect on score (ANOVA: P-value ≥ 0.05). However ‘number of significant kinases’ did have a significant effect on score (ANOVA: P-value < 0.0001; Pearson Correlation: r = 0.3; P-value < 0.0001) (Figure 4). Lastly, the average score of each target kinase (when it was selected in the course of the random search) was calculated (Figure 5). Of the target kinases that were correctly predicted, average score varied significantly (ANOVA: P-value < 0.0001). The kinases with the top five average scores were: *CDK4*, *HIPK2*, *CDK2*, *GSK3B*, and *CHUK*.

**Summary**:

* Of each step in X2K, only the PPI databases had a significant impact on the score (-log(P-value)).
* Stelzl, PPID, IntAct, MINT and BioGRID were the best PPI databases according to the random search.
* ‘Number of significant kinases’ is correlated with higher score, although the complex nature of X2K makes the reasons behind this unclear.
* There is considerable variation in how well certain kinases are predicted (although it is unclear whether this stems from biases in X2K, or the composition of the GEO validation database).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **SS** | **DF** | **F-value** | **P-value** | **Sig.** |
| TF Database | 575 | 7 | 5 | < 0.0001 | \*\*\*\* |
| PPI Database | 2188 | 16 | 9 | < 0.0001 | \*\*\*\* |
| Kinase Database | 101 | 7 | 1 | ≥ 0.05 | non-sig |

**Table 1**. Summary table of the ANOVA results for each database type (TF, PPI, kinase) vs. -log(P-value) of all correctly predicted target kinases. SS =Sum of Squares; DF = Degrees of Freedom; Sig. = Significance (non-sig. = non-significant/p≥ 0.05; \*= p<0.05; \*\*= p<0.01; \*\*\*p<0.001).



**Figure 1**. Transcription factor databases plotted against -log(P-value) of correctly predicted kinases across all individuals (parameter combinations ) in the random search. Databases are in order of descending -log(P-value). Individuals that did not recover the correct kinases were filtered out for visualization purposes.

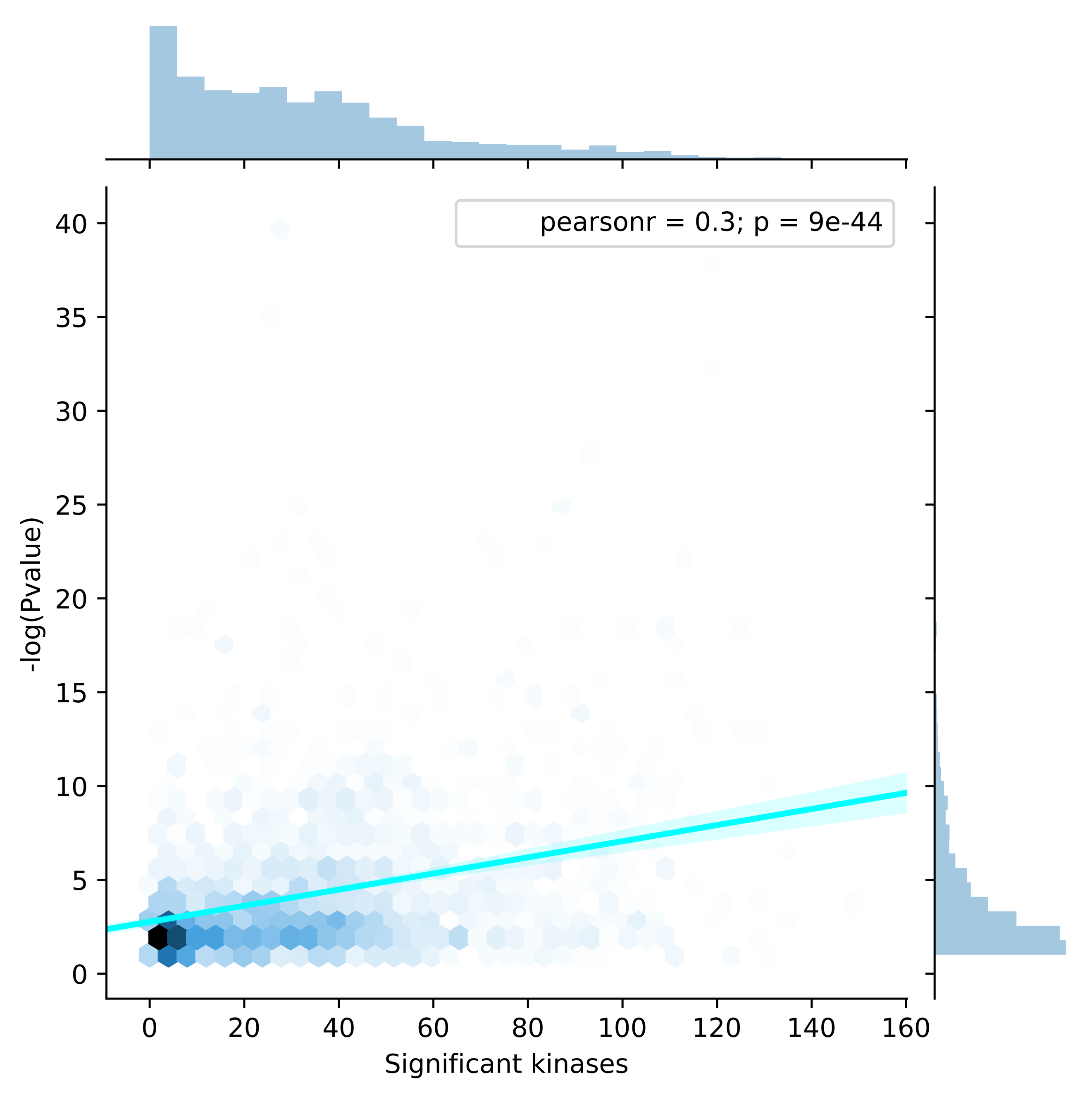


**Figure 2**. PPI databases plotted against -log(P-value) of correctly predicted kinases

across all individuals (parameter combinations ) in the random search. Databases are in order of descending -log(P-value). Individuals that did not recover the correct kinases were filtered out for visualization purposes.



**Figure 3**. Kinase databases plotted against -log(P-value) of correctly predicted kinases across all individuals (parameter combinations) the random search. Databases are in order of descending -log(Pvalue). Individuals that did not recover the correct kinases were filtered out for visualization purposes.



**Figure 4**. Number of significant kinases plotted against -log(P-value) of correctly predicted kinases across all individuals (parameter combinations). A linear regression as used to plot a best fit line (in cyan; Pearson r = 0.3; 2.3e-10).



**Figure 5**. Target kinases (from validation data experiments) plotted against -log(P-values) of correctly predicted kinases across all individuals (parameter combinations). Kinases are in order of descending -log(P-value). Individuals that did not recover the correct kinases were filtered out for visualization purposes.



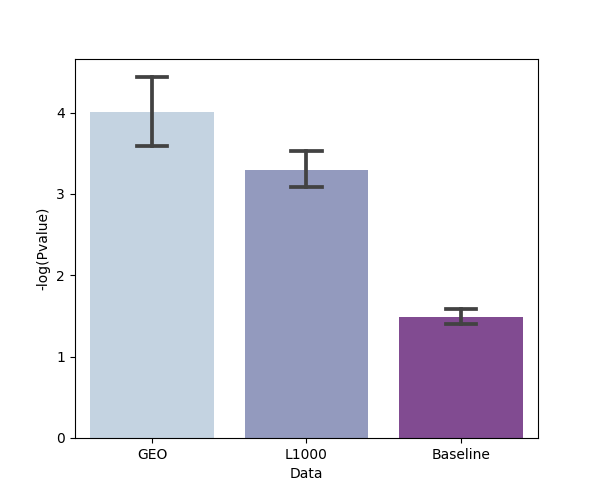
**Figure 6**. For each TF database, the % of instances in the random search that (when selected) X2K was able to correctly recover the target kinase.



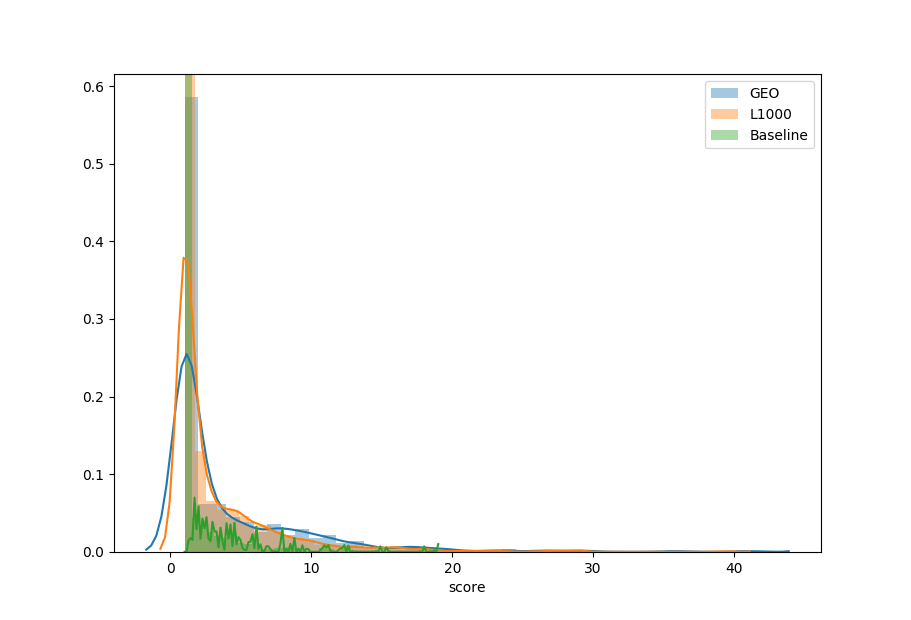
**Figure 7**. For each PPI database, the % of instances in the random search that (when selected) X2K was able to correctly recover the target kinase.



**Figure 8**. For each kinase database, the % of instances in the random search that (when selected) X2K was able to correctly recover the target kinase.



**Figure 9**. Barplot of the scores resulting from X2K runs using optimized parameters on GEO kinase perturbation data (n = 570 experiments), optimized parameters on L1000 kinase perturbation data (n = 1400 experiments), and non-optimized parameters using random search on the GEO data (i.e. Baseline; n = 570).



**Figure 9**. Normalized histogram of the scores resulting from X2K runs using optimized parameters on GEO kinase perturbation data (n = 570 experiments), optimized parameters on L1000 kinase perturbation data (n = 1400 experiments), and non-optimized parameters using random search on the GEO data (i.e. Baseline; n = 570).