Calculation of the Upstream Kinase Analysis (UKA) scores

Peptides on the PamChip can be phosphorylated by a number of .Kinases. Each .Kinase can phosphorylate multiple Peptides on the PamChip. Information on how we calculate prediction scores is explained. Contact us for further details on the algorithms used.

1. Peptide Difference Statistic

The Peptide Difference Statistic (Peptide Statistic) is calculated for each Peptide comparing two conditions, using the $\log_2$ .Kinase activity profiles measured on the chips. The Peptide Statistic is the Signal to Noise Ratio (SNR) of the difference between the $\log_2$ .Kinase activity profiles.

$$p = \frac{\overline{x_T} - \overline{x_{C=0}}}{\sqrt{\sigma_T^2 + \sigma_C^2}}$$

where p = Peptide Statistic, $\bar{x}$ = mean $\log_2$ .Kinase activity of T (Test) or C (Control) conditions, $\sqrt{\sigma^2}$ = standard deviation of .Kinase activity of the two conditions.

A similar value that describes the Peptide difference is the Peptide Change:

$$p_{change} = \overline{x_T} - \overline{x_{C=0}}$$

This is used to calculate the Kinase Change (see later).

2. Upstream Kinase (UK) to Peptides ranking

Peptides on the PamChip represent phosphosites for tyrosine (PTK PamChip) or serine-threonine .Kinases (STK PamChip). Using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) Peptide sequences on the PamChips were matched to phosphosites in the human proteome. 6 post translational modification (PTM) databases were then queried for .Kinases that are verified *(in vitro/ in vivo* evidence) and/or predicted *(in silico,* PhosphoNET) to phosphorylate these sites (Hence the term Upstream Kinase)

This results in a binarized Peptide-to-.Kinase matrix, U, where zeroes represent no connection between Peptides and Upstream Kinases (UKs) and ones represent a connection.

Matrix U

peptides \ *kinases*			
	1	0	1
	0	1	1

There is a connection (a Peptide is enriched with an UK) if:

- the Peptide has a match to a phosphosite for which the UK is found in an in *vivo / in vitro* database (Kinase rank designated as 0).
- the Peptide has a match to a phosphosite for which the UK has a PhosphoNET (*in silico* database) 'Kinase Predictor V2 Score' > 300 and a rank less than or equal to a cut-off (see in the next section, at a default of Rank 12).

At the end of this step, each identified .Kinase has a list of Peptides it is predicted to phosphorylate, and each Kinase to Peptide relationship has a rank, either 0 from the verified databases or 1 to 12 from the in silico database.

3. For each .Kinase, different size of Peptide sets are considered

In each analysis, multiple U matrices are calculated based on varying the Kinase rank cutoff. Thus, for each .Kinase, multiple Peptide sets are considered – by decreasing the stringency of the cutoff, the size of the Peptide set increases. The default set includes the Peptides of ranks 0 to 4 and subsequent Peptide sets are formed by including the lower ranks from 5 to 12, one-by-one/ additively.

4. The Kinase Statistic and the Kinase Change

The Median Kinase Statistic represents the direction of effect: the change in .Kinase activity in a Test condition (T) compared to a Control condition (C). < 0 means inhibition, > 0 means activation in T versus C. The Kinase Statistic is the median of the Peptide Statistics of the set of Peptides that a .Kinase can phosphorylate. Thus, the Kinase Statistic represents the log fold change scaled by the noise.

For each U matrix (for each Peptide set per .Kinase), the normalized Kinase Statistic is calculated:

$$\boldsymbol{s = b * (U^T p)}$$

$$\boldsymbol{b_i = 1/a_i}, \text{ where}$$

$\boldsymbol{s}$: the Kinase Statistic is the weighted average of $\boldsymbol{p}$, the Peptide difference statistic. $\boldsymbol{b}$ is a normalization vector, $\boldsymbol{a_i}$ is the number of nonzero elements in column i of U (the number of Peptides that a .Kinase phosphorylates).

An example calculation:

$$\underset{\text{kinases}}{\begin{matrix}\\ \\ \\ \end{matrix}}\overset{\begin{matrix}U^T \\ \text{peptides}\end{matrix}}{\begin{pmatrix}0 & 1\\ 1 & 0\\ 1 & 1\end{pmatrix}} \times \overset{\begin{matrix}\text{peptide difference} \\ \text{statistic (p)}\end{matrix}}{\begin{pmatrix}0.5\\ 1.5\end{pmatrix}} = \overset{\begin{matrix}\text{Kinase} \\ \text{statistic}\end{matrix}}{\begin{pmatrix}1.5\\ 0.5\\ 2\end{pmatrix}} \Longrightarrow \begin{matrix}\text{Normalized kinase statistic} \\ 1.5/1 = 1.5 \\ 0.5/1 = 0.5 \\ 2/2 = 1.0\end{matrix}$$

Formulated in a different way: for .Kinase i,

$$s_i = \frac{1}{a_i} * \sum_{j=1}^{k} p_j$$

For each Peptide set considered for a .Kinase, the Kinase Statistic is calculated, and in the results, the Median Kinase Statistic across different Peptide sets is given.

The **Kinase Change** is calculated similarly as the Kinase Statistic, but using the p_{change} values:

$$\boldsymbol{s_{MKC} = b * (U^T p_{change})}$$

The Kinase Statistic, in contrast to Kinase Change, is a value that also describes the significance of the change, i.e. whether the change is higher than the noise.

5. Significance Score

The significance of the Kinase Statistic is based on a permutation test where the samples are permuted. The Kinase Statistic is recalculated for each permutation. The Significance Score is based on the difference between the Kinase Statistic of the actual sample versus that of the permuted samples. A high Significance Score means a high probability that the .Kinase is differentially active between Test and Control.

6. Specificity Score

The specificity of the Kinase Statistic is based on a permutation test, where the Peptides are permuted. The Kinase Statistic is recalculated for each permutation. The Specificity Score is based on the difference between the Kinase Statistic of the actual sample versus that of the permuted samples. A high Specificity Score means a high probability that the observed effect could not have been obtained by a random set of Peptides.

7. Final Score

The Final Score is the sum of the Significance Score and the Specificity Score (Figure 1). The Final Score is used to rank the .Kinases for putative involvement in the observed experimental differences. The default threshold is >1.3.

8. UKA Score Table

A score table is generated as output. Mean and median scores are calculated across the different Peptide sets considered for a Kinase. For further analysis, the Median Kinase Statistic and Median Final Score are used.

Figure 1. Summary of PamGene's UKA Functional Class Scoring Tool