PamGene Supplement:

Methods and Data analysis

*Peter Linders for Odile Filhol-Cochet*

January 25, 2022

Table of Contents

[1. Introduction 2](#_Toc94016370)

[1.1. PamChip technology and assay principle 2](#_Toc94016371)

[2. Data analysis 4](#_Toc94016372)

[2.1. Data analysis overview 4](#_Toc94016373)

[2.2. Basic processing 5](#_Toc94016374)

[2.2.1. Image Analysis and signal integration 5](#_Toc94016375)

[2.2.2. Quality Control 5](#_Toc94016376)

[2.2.3. Visual assessment of Overall signal 7](#_Toc94016377)

[2.3. Per phosphosite analysis (Statistics) 9](#_Toc94016378)

[2.3.1. Differential statistical analysis methods: “Test condition (s)” versus “Control” 9](#_Toc94016379)

[2.3.2. MTvC results 9](#_Toc94016380)

[2.3.3. T-Test results 11](#_Toc94016381)

[2.4. Kinase Analysis 12](#_Toc94016382)

[2.4.1. Kinase interpretation (Differential) 12](#_Toc94016383)

[2.5. Biological interpretation 18](#_Toc94016384)

[2.5.1. Pathway Analysis (GProfiler, Gene Ontology Biological Processes) 19](#_Toc94016385)

[3. Glossary 20](#_Toc94016386)

# 

# Introduction

This supplement gives a detailed explanation of the methods used in the main study report.

With PamGene's kinase activity profiling platform, the activity of kinases in a wide range of cells and tissues can be measured in real time. We use our proprietary 3D peptide microarray technology (PamChip® and PamStation®) which offers a multiplex method for global kinase activity profiling. The assay is very sensitive, requiring only a small amount of lysate to measure the activity of kinases in various samples including cell lines, xenografts and human tissues. Lysates obtained from a few thousand cells can suffice to obtain a kinome profile of the multiple kinases present in these samples. This is accomplished by incubating the sample lysates across peptide substrates immobilized on the 3D surface of the PamChip® microarray (196 protein tyrosine kinase (PTK) or 144 serine/threonine kinase (STK), hereafter referred to as phosphosites). Kinases present in the lysates will phosphorylate the phosphosites, which are visualized using fluorescently labelled antibodies.

## PamChip technology and assay principle

PamGene's microarray assay for kinase activity profiling is based on measuring phosphorylation of phosphosites by protein kinases. The PamChip® 4 consumable consists of 4 identical arrays, each array containing 144 (STK) or 196 (PTK) phosphosites immobilized on a porous ceramic membrane (Figure 1). These phosphosites are encoded in 13 amino acid long peptides which derived from literature or computational predictions. with the phosphorylation of these phosphosites is then used to predict one or multiple upstream kinases (Protein tyrosine kinases for the PTK PamChip® and Serine threonine kinases for the STK PamChip®). Fluorescently labelled anti-phospho-antibodies are used to detect phosphorylation activity of kinases present in the sample (Figure 1).

During the assay, the sample is pumped through the porous membrane, allowing for shorter assay times. When the solution is underneath the array, images of each array are taken at several exposure times by a camera in the workstation (Figure 1). Images are later used by the BioNavigator® software to calculate signal values for each phosphosite. The data workflow consisting of image quantification, quality control, statistical analysis, visualization and interpretation is performed using the BioNavigator® software.

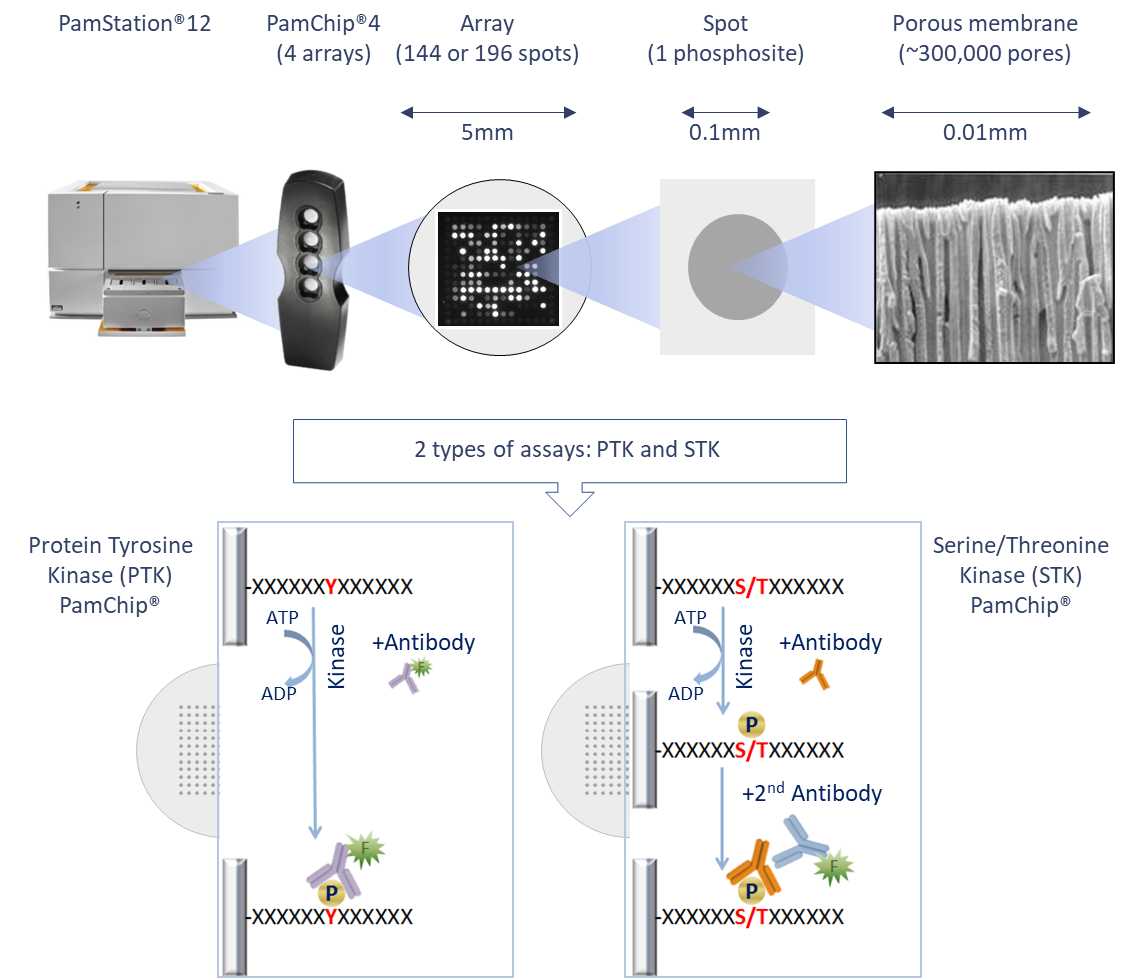


Figure . PamGene technology outline and PamChip assay principle

# Data analysis

## Data analysis overview

The general workflow for data analysis is shown below (Figure 2), which was used to derive the results described in the Report and this Supplement, with details following.

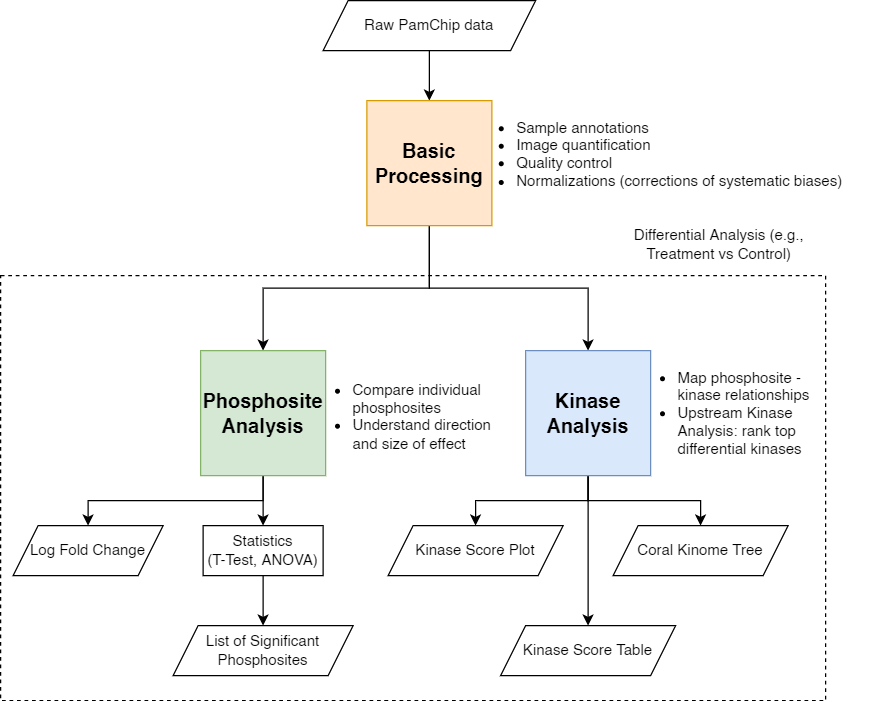


Figure . Data analysis workflow. Parallelograms indicate in- and outputs. Basic Processing (orange box) refers to Section 2.2 of this supplement. Phosphosite Analysis (green box) refers to Section 2.3. Kinase Analysis (blue box) refers to Section 2.4. The output of Phosphosite Analysis and Kinase Analysis can be used for further biological interpretation (see also Section 2.5).

## Basic processing

### Image Analysis and signal integration

**Image analysis**

The image analysis is performed on each image. Once a grid is constructed for the image, various unique values (quantitation types) are calculated. The main value considered as the “signal” is called the Median\_SigmBg (median signal minus local background). Signal\_Saturation (fraction of saturated pixels) values are used to remove saturated spots during signal integration.

**Exposure Time Integration**

Low exposure times are useful to capture high intensity signals, and likewise high exposure times are useful to capture low intensity signals. Therefore, the same image is acquired at multiple exposure times, and then Median\_SigmBg values are integrated to a single value, called the S100 signal. Since saturated spots (5% saturation) are excluded in the integration, the dynamic range of the measurements is increased (the ratio of the highest and lowest signal that can be measured).

### Quality Control

The quality of the data was assessed using criteria shown in Table 1. The QC results for the data in this study are shown in Table 2.

Three criteria were checked using a flag system is used to indicate a quality level,

* Phosphosite signal strength (i.e., the percentile 0.95 of S100 AU (arbitrary units).
* Phosphosite number control (i.e., how many phosphosites passed quality control).
* Sample quality control (i.e., Replicate CV (coefficient of variance); An example of technical replicate is 1 sample lysate on 3+ arrays; An example of biological replicate is 1 sample lysate per array).

Table . Flag system used to assess data quality

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Level** | **Flag1** | **Signal strength (AU)** | **QC-passed phosphosites (PTK) of 195** | **QC-passed phosphosites (STK) of 142** | **Technical CV** | **Biological CV** |
| Good |  | 0-4,000 | >123 | >90 | <20% | <30% |
| Fair |  | 0-2,000 | 78-123 | 56-90 | 20%-30% | 30%-40% |
| Poor |  | 0-1,000 | <78 | <56 | >30% | >40% |

1 The QC flag is based on the 3 individual criteria. All criteria must be met to get a green QC flag.

**Number of QC-passed phosphosites (PTK)**

For the PTK Assay there is a kinetic readout, with images obtained every 5 minutes during the assay. The phosphorylation kinetics for each phosphosite (each spot) is analyzed by plotting S100 values versus the Time (cycles, in minutes). Only phosphosites that show kinetics (increase of signal in time) on at least 25 % of the arrays are included in the downstream analysis (kinetic fraction present > 0.25).

**Number of QC-passed phosphosites (STK)**

Phosphosite QC selection for the STK assay is performed by carrying out a nominal CV estimation. Phosphosites with a CV lower than 50% over all measured arrays are included in the downstream analysis. Values after wash (Cycle 124; for the QC phosphosites) are log2 transformed and values below 0 are removed.

Table . Flag system used to assess the quality of the data in the study

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Assay Type** | **Flag** | **Signal strength (AU)** | **QC passed phosphosites** | **Technical CV** | **Biological CV** |
| **PTK** |  | 848 | 130 / 192 | n.a..1 | 11-30% |
| **STK** |  | 1186 | 92 / 144 | n.a..1 | 14-18% |

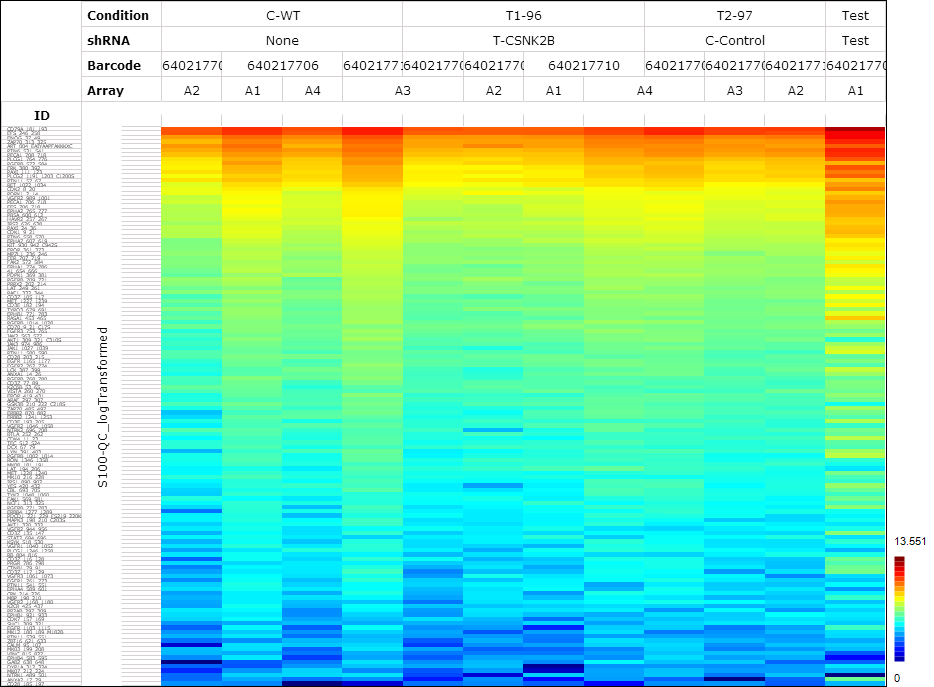
1 No technical replicates were run in this assay; therefore, the technical variance was not applicable

### Visual assessment of Overall signal

The basic processing steps described above result in Log2-transformed signal values for each peptide (rows) and each array (columns). Signals before and after normalization are shown below.

**Heatmap log s100 (PTK) Overview (before normalization)**

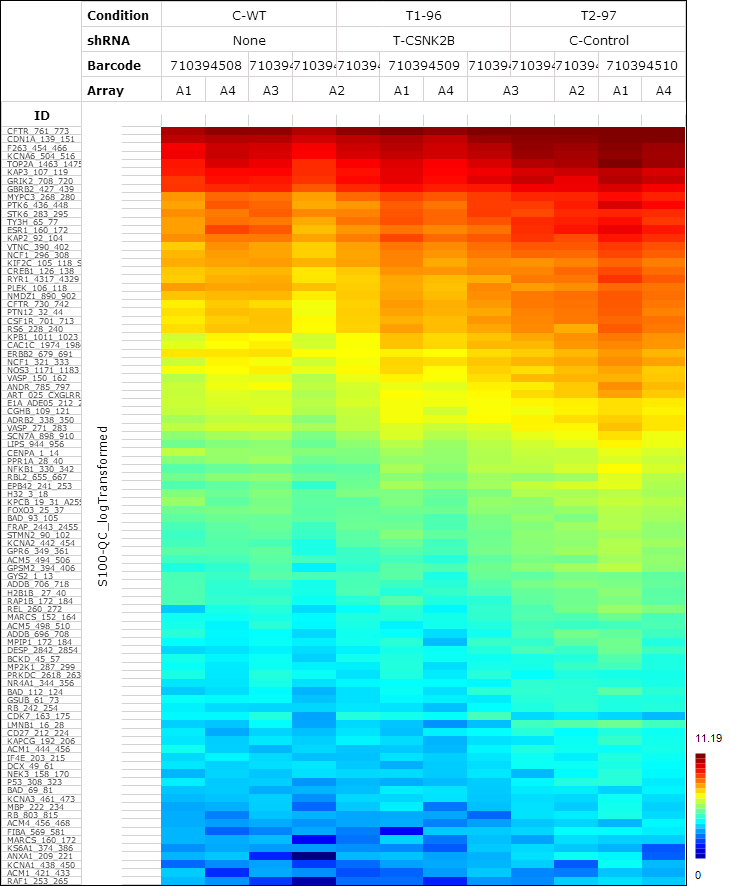
The heatmap visualization below provides an overview of samples and measurements. This view helps to indicate any possible trends and outliers.



##### The heatmap shows log2 values of the integrated signal. Each row is a phosphosite and each column is a PamChip array. Rows are sorted by row mean and only include phosphosites which passed the QC..

**Heatmap log s100 (STK) Overview**

The heatmap visualization below provides an overview of samples and measurements. This view helps to indicate any possible trends and outliers.



##### The heatmap shows log2 values of the integrated signal. Each row is a phosphosite and each column is a PamChip array. Rows are sorted by row mean and only include phosphosites which passed the QC.

## Per phosphosite analysis (Statistics)

### Differential statistical analysis methods: “Test condition (s)” versus “Control”

To identify significant differences between the conditions at the phosphosite level, one of the following statistical tests are used:

* Multiple Test conditions versus Control (MTvC): To compare 2 or more Test conditions versus Control. An ANOVA is performed for significant effects anywhere between the Test conditions (including control). A Dunnett’s post-hoc test is then performed for multiple comparisons to find significant effects between each Test condition.
* T-test (TT): To compare 1 Test condition versus Control: Two-sided Student’s T-test (TT) that are paired or unpaired are performed.

The following plots show the results from the statistical tests:

* Volcano plot, useful for a Test condition effect overview (A volcano plot is used to give an overview of effect size, direction and significance)
* Significant phosphosites (LFC or Scaled heatmaps)

Results from the above tests can be provided upon request, in Excel format

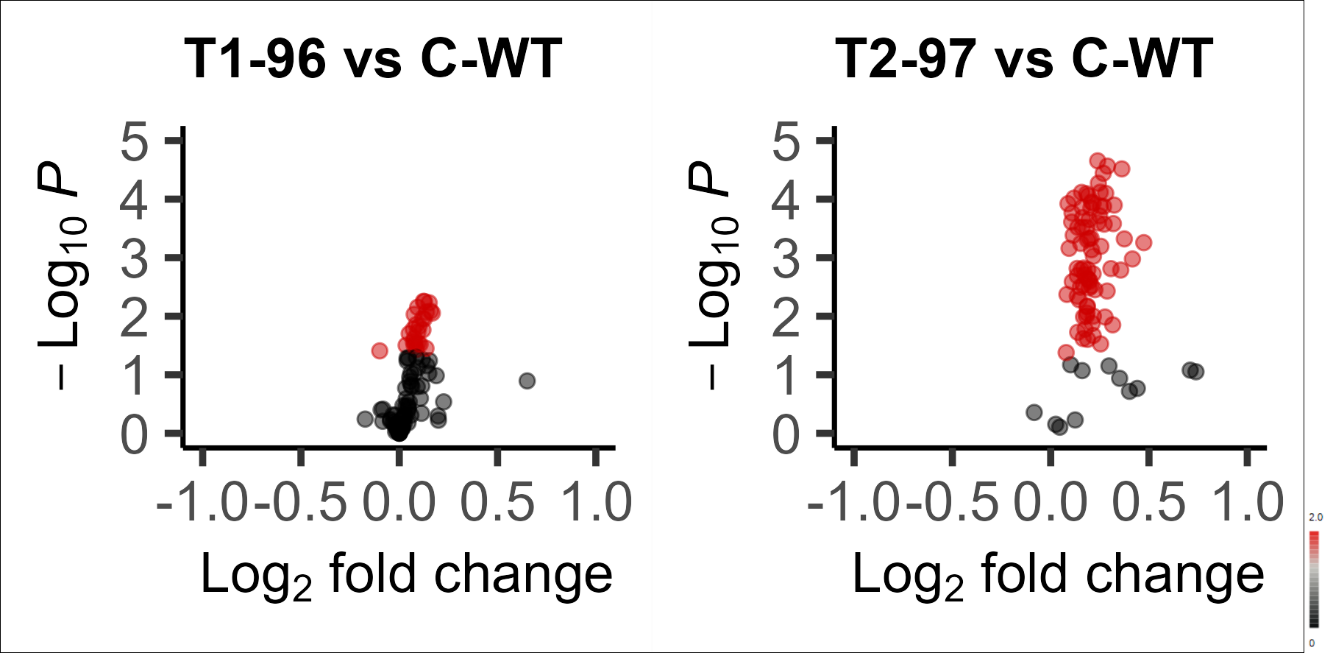
### MTvC results

**Volcano Plot PTK**



##### The volcano plot visualizes the result of the tests by plotting – for each test – the effect size (x-axis, LFC or delta) versus significance (y-axis, -log10(p value)) of the test. Red spots are phosphosites that show significant difference compared to control (p<0.05).

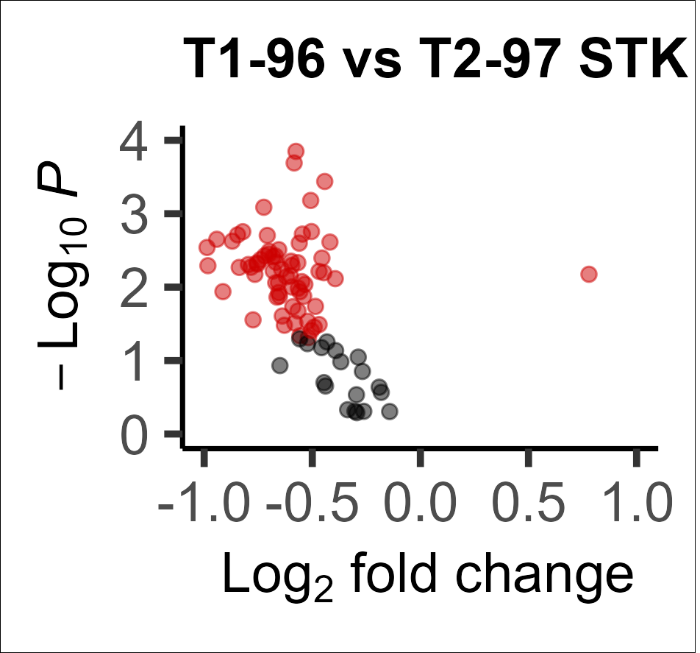
**Volcano Plot STK**

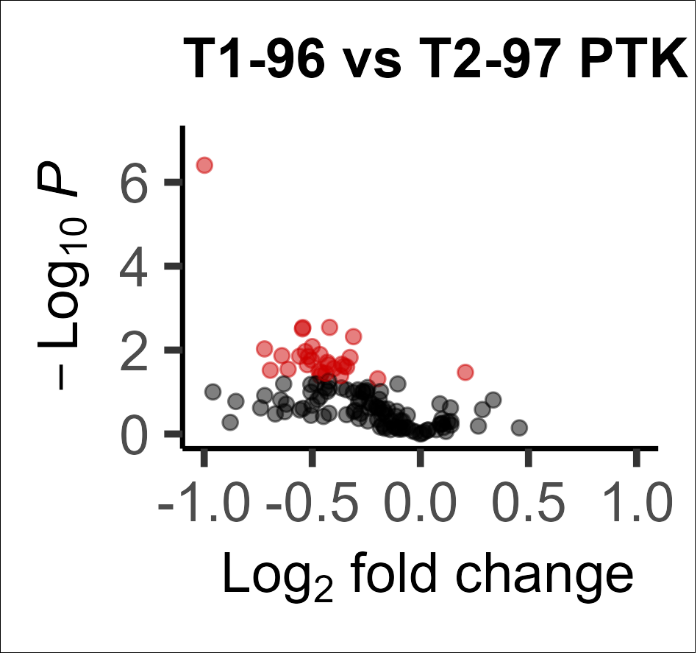


##### The volcano plot visualizes the result of the tests by plotting – for each test – the effect size (x-axis, LFC or delta) versus significance (y-axis, -log10(p value)) of the test. Red spots are phosphosites that show significant difference compared to control (p<0.05).

### T-Test results

**Volcano Plots T1-96 vs T2-97**





##### The volcano plot visualizes the result of the tests by plotting – for each test – the effect size (x-axis, LFC or delta) versus significance (y-axis, -log10(p value)) of the test. Red spots are phosphosites that show significant difference compared to control (p<0.05).

## Kinase Analysis

### Differential upstream kinase analysis methods: “Test condition” versus “Control”

The **Upstream Kinase Analysis (UKA)** algorithm is used to predict differential kinase activity in the test condition compared to the control. The UKA uses knowledge from publicly available databases that specify kinase-to-substrate relationships. Therefore, the interpretation of the derived kinases from the UKA is highly dependent on the contents of these databases. Ultimately, the results from the UKA can be used to generate hypotheses and the selected kinases need to be further validated using different approaches. Whether they make sense in a biological context is dependent on several factors (e.g., the model used in the experiments) and requires further consideration.

The following plots are generated:

* **Kinase Score Table**, useful to identify the top kinases
* **Kinase Score Plot**, useful to identify the specificity of kinases (see below)
* **Kinome Tree**, useful to group the kinases into sequence families (<http://phanstiel-lab.med.unc.edu/CORAL/>). Additional information and tutorial are available on request.

The table and plots visualize the results from the UKA, which uses functional class scoring (similar to gene set enrichment analysis) to determine which kinases are active by testing which sets of phosphosites have significant effects. UKA returns several different values:

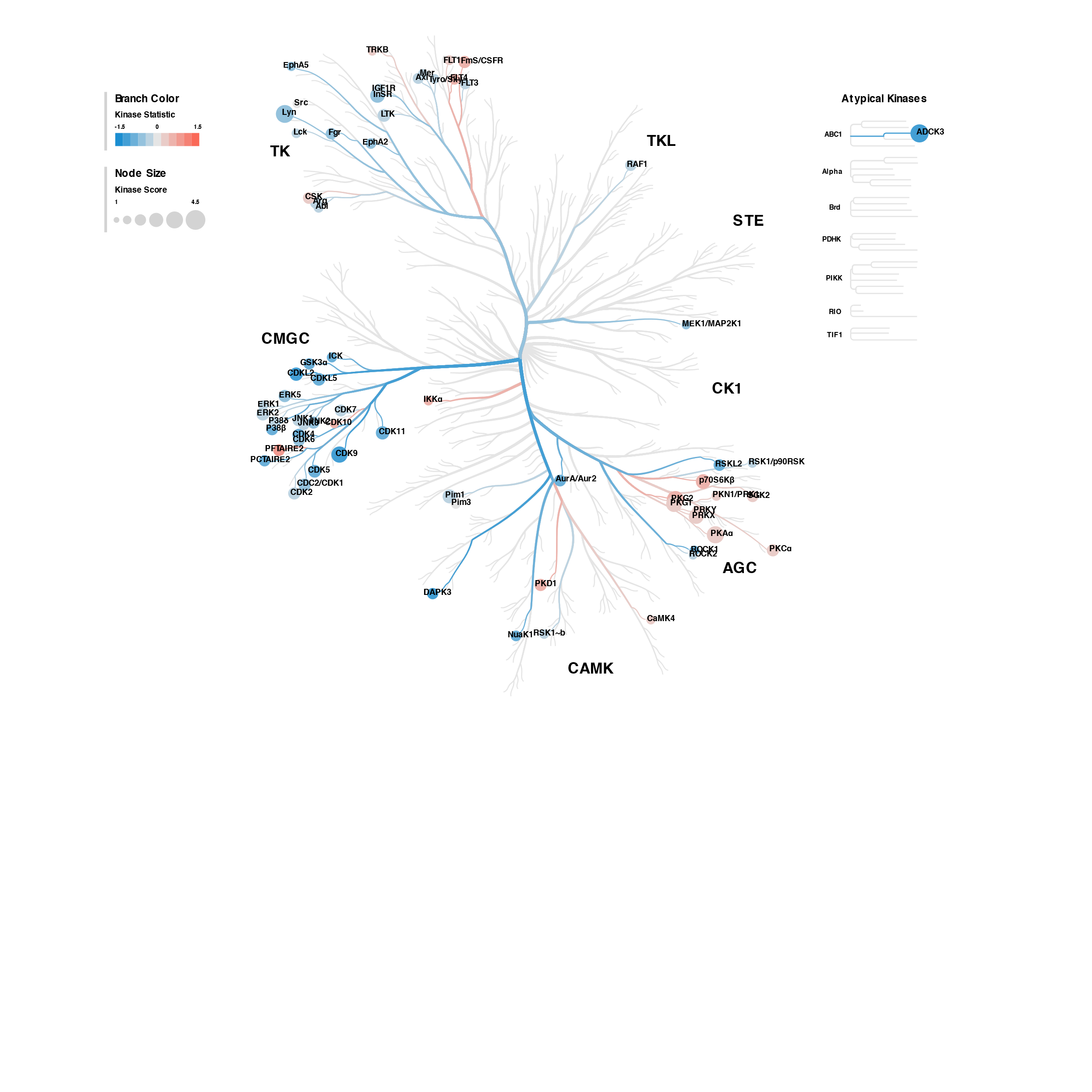
* Specificity score: represents the specificity of the change in kinase activity. A higher specificity score means a lower chance that the observed effect could have been obtained by a random set of peptides.
* Kinase statistic: represents the change in kinase activity, < 0 = inhibition, > 0 = activation.
* Final score: combines the significance of the result with the specificity, kinases are ordered by this quantitation. Kinases with a high final score have a higher probability of being differentially active within the test condition.

A general guideline below can be used to infer biological interpretation from the UKA results. PamGene can support and discuss these approaches.

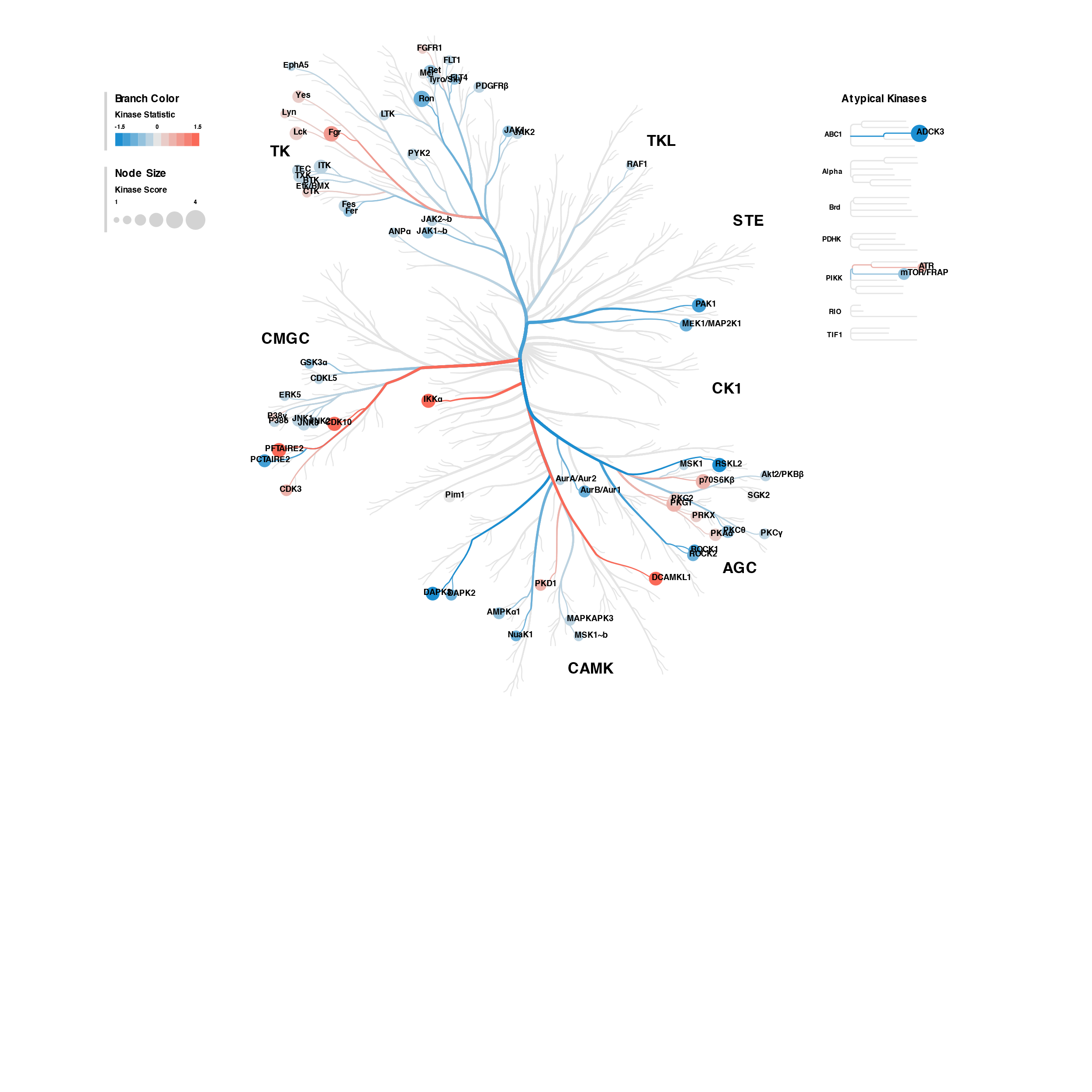
1. Use the UKA tool to give a ranked predicted list of kinases
2. Select top kinase(s) based on:
   1. Ranking order
   2. Threshold cut-off (subjective; guideline is to use Median Final score > 1.2)
   3. Known biological information
3. Further validate the selected kinase based on:
   1. Biological context inferred from current knowledgebase and literature
   2. Other platforms and technologies

More details of this algorithm are available on request. Results are provided as Excel (txt) files.

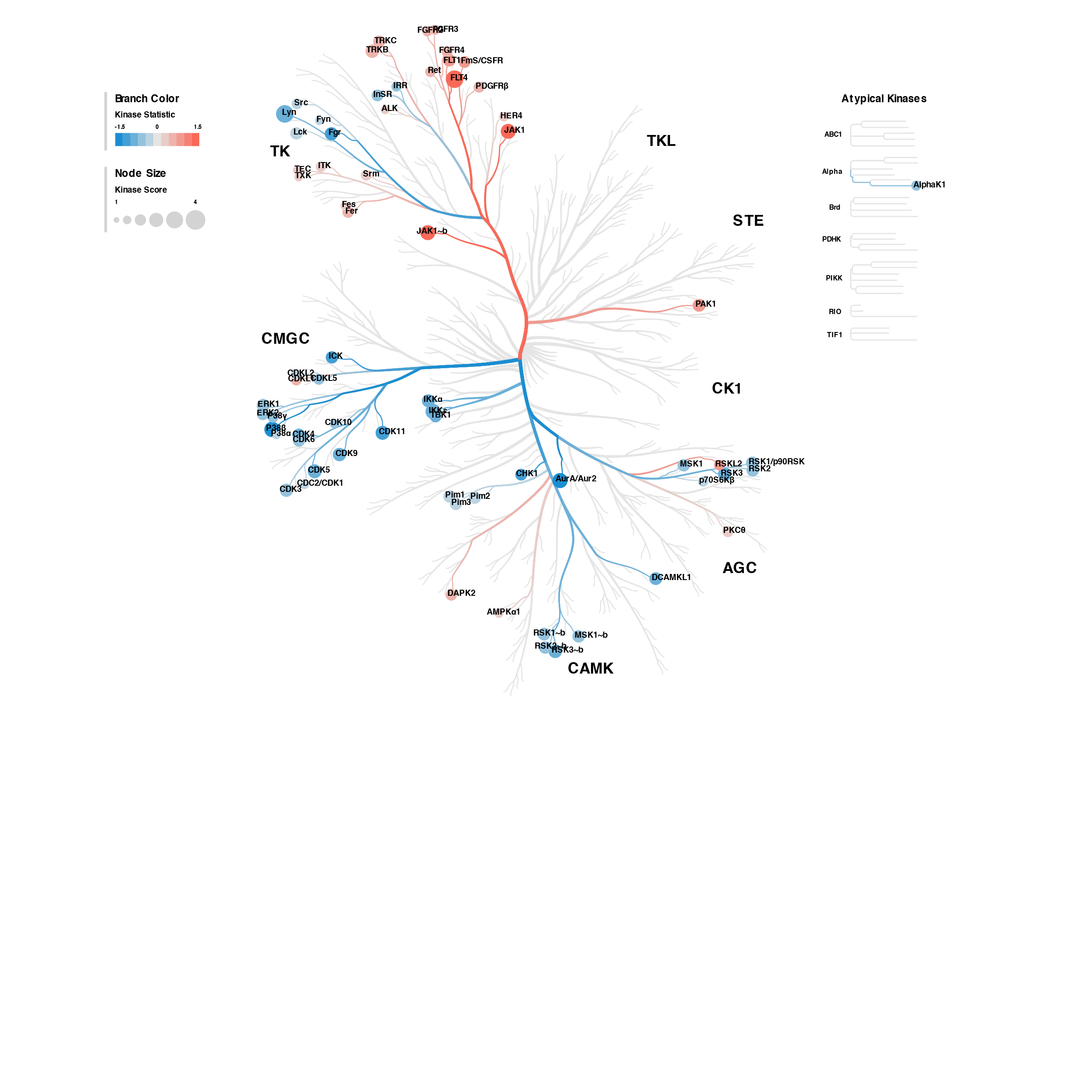
**Kinome trees (PTK + STK) T1-96 v Control-WT**

****

**Kinome trees (PTK + STK) T2-97 v Control-WT**

****

**Kinome trees (PTK + STK) T1-96 v T2-97**

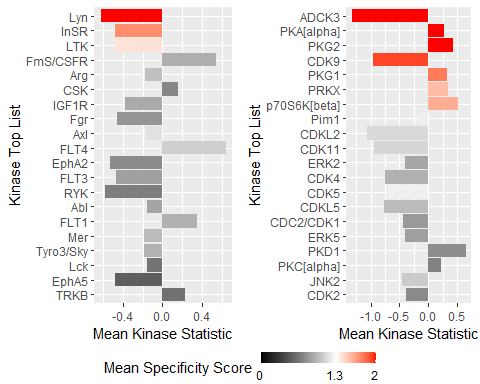
****

**Kinome trees:** Top predicted differentially active kinases are represented on the phylogenetic tree of the human protein kinase family. The size of the dots indicates the total kinase score and the color denotes the kinase statistic. Kinases that are less active than the average kinase activity are shown in blue, kinases that are more active are shown in red.

**Kinase Score Table (PTK + STK) of T1-96 vs Control-WT**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **PTK** | | | **STK** | | |
| **Rank** | **Kinase Name** | **Kinase Score** | **Kinase Statistic** | **Kinase Name** | **Kinase Score** | **Kinase Statistic** |
| 1 | Lyn | 4.14 | -0.63 | ADCK3 | 4.22 | -1.32 |
| 2 | InSR | 3.11 | -0.48 | PKA[alpha] | 3.92 | 0.28 |
| 3 | LTK | 2.53 | -0.48 | PKG2 | 3.92 | 0.43 |
| 4 | FmS/CSFR | 2.32 | 0.54 | CDK9 | 3.62 | -0.96 |
| 5 | Arg | 2.19 | -0.17 | PKG1 | 3.49 | 0.32 |
| 6 | CSK | 2.16 | 0.16 | PRKX | 3.16 | 0.34 |
| 7 | IGF1R | 2.06 | -0.38 | p70S6K[beta] | 3.11 | 0.52 |
| 8 | Fgr | 2.04 | -0.46 | Pim1 | 2.88 | -0.16 |
| 9 | Axl | 1.92 | -0.17 | CDKL2 | 2.69 | -1.06 |
| 10 | FLT4 | 1.90 | 0.64 | CDK11 | 2.69 | -0.94 |
| 11 | EphA2 | 1.81 | -0.53 | ERK2 | 2.53 | -0.41 |
| 12 | FLT3 | 1.70 | -0.47 | CDK4 | 2.50 | -0.74 |
| 13 | RYK | 1.63 | -0.58 | CDK5 | 2.49 | -0.69 |
| 14 | Abl | 1.59 | -0.16 | CDKL5 | 2.47 | -0.77 |
| 15 | FLT1 | 1.57 | 0.35 | CDC2/CDK1 | 2.37 | -0.44 |
| 16 | Mer | 1.56 | -0.19 | ERK5 | 2.35 | -0.40 |
| 17 | Tyro3/Sky | 1.49 | -0.19 | PKD1 | 2.32 | 0.66 |
| 18 | Lck | 1.46 | -0.16 | PKC[alpha] | 2.31 | 0.22 |
| 19 | EphA5 | 1.35 | -0.48 | JNK2 | 2.31 | -0.46 |
| 20 | TRKB | 1.27 | 0.22 | CDK2 | 2.30 | -0.39 |

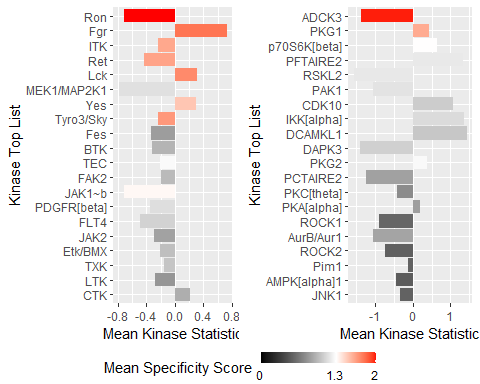
**Kinase Score Plot (PTK left, STK right) of T1-96 vs Control-WT**



**Kinase Score Table (PTK + STK) of T2-97 vs Control-WT**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **PTK** | | | **STK** | | |
| **Rank** | **Kinase Name** | **Kinase Score** | **Kinase Statistic** | **Kinase Name** | **Kinase Score** | **Kinase Statistic** |
| 1 | Ron | 3.34 | -0.71 | ADCK3 | 3.62 | -1.38 |
| 2 | Fgr | 3.08 | 0.73 | PKG1 | 3.08 | 0.43 |
| 3 | ITK | 2.72 | -0.24 | p70S6K[beta] | 2.92 | 0.64 |
| 4 | Ret | 2.68 | -0.43 | PFTAIRE2 | 2.77 | 1.33 |
| 5 | Lck | 2.55 | 0.31 | RSKL2 | 2.75 | -1.56 |
| 6 | MEK1/MAP2K1 | 2.41 | -0.79 | PAK1 | 2.74 | -1.05 |
| 7 | Yes | 2.32 | 0.30 | CDK10 | 2.74 | 1.06 |
| 8 | Tyro3/Sky | 2.28 | -0.24 | IKK[alpha] | 2.71 | 1.36 |
| 9 | Fes | 2.21 | -0.34 | DCAMKL1 | 2.62 | 1.43 |
| 10 | BTK | 2.10 | -0.32 | DAPK3 | 2.61 | -1.40 |
| 11 | TEC | 2.06 | -0.21 | PKG2 | 2.55 | 0.37 |
| 12 | FAK2 | 2.01 | -0.20 | PCTAIRE2 | 2.38 | -1.22 |
| 13 | JAK1~b | 1.98 | -0.71 | PKC[theta] | 2.35 | -0.41 |
| 14 | PDGFR[beta] | 1.94 | -0.35 | PKA[alpha] | 2.25 | 0.21 |
| 15 | FLT4 | 1.84 | -0.49 | ROCK1 | 2.17 | -0.88 |
| 16 | JAK2 | 1.77 | -0.28 | AurB/Aur1 | 2.10 | -1.05 |
| 17 | Etk/BMX | 1.76 | -0.20 | ROCK2 | 2.09 | -0.74 |
| 18 | TXK | 1.73 | -0.16 | Pim1 | 2.09 | -0.13 |
| 19 | LTK | 1.58 | -0.28 | AMPK[alpha]1 | 2.06 | -0.44 |
| 20 | CTK | 1.48 | 0.22 | JNK1 | 2.06 | -0.33 |

**Kinase Score Plot (PTK left, STK right) of T2-97 vs Control-WT**



**Kinase Score Table (PTK + STK) of T1-96 vs T2-97**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **PTK** | | | **STK** | | |
| **Rank** | **Kinase Name** | **Kinase Score** | **Kinase Statistic** | **Kinase Name** | **Kinase Score** | **Kinase Statistic** |
| 1 | Lyn | 3.70 | -0.83 | p38[beta] | 3.12 | -1.50 |
| 2 | FLT4 | 3.70 | 1.38 | IKK[epsilon] | 3.02 | -1.16 |
| 3 | JAK1~b | 2.96 | 1.27 | AurA/Aur2 | 3.01 | -1.24 |
| 4 | TRKB | 2.52 | 0.54 | CDK11 | 2.67 | -1.17 |
| 5 | Fgr | 2.44 | -0.99 | CDK5 | 2.67 | -0.79 |
| 6 | FLT1 | 2.32 | 0.77 | RSK1/p90RSK | 2.53 | -0.91 |
| 7 | Lck | 2.22 | -0.36 | CDK9 | 2.51 | -0.81 |
| 8 | TRKC | 2.13 | 0.46 | CDK3 | 2.46 | -0.67 |
| 9 | FmS/CSFR | 2.07 | 0.82 | IKK[alpha] | 2.44 | -0.99 |
| 10 | CHK1 | 2.06 | -1.15 | CDK4 | 2.42 | -0.80 |
| 11 | InSR | 2.04 | -0.51 | ERK1 | 2.39 | -0.50 |
| 12 | Fer | 1.96 | 0.61 | PAK1 | 2.36 | 0.70 |
| 13 | PDGFR[beta] | 1.85 | 0.53 | RSK3 | 2.35 | -0.57 |
| 14 | FGFR4 | 1.84 | 0.35 | DCAMKL1 | 2.33 | -0.78 |
| 15 | Srm | 1.80 | 0.26 | Pim1 | 2.25 | -0.20 |
| 16 | Src | 1.78 | -0.22 | RSK2 | 2.24 | -0.51 |
| 17 | Ret | 1.69 | 0.45 | MSK1 | 2.23 | -0.53 |
| 18 | ITK | 1.63 | 0.29 | CDK6 | 2.20 | -0.62 |
| 19 | FGFR2 | 1.61 | 0.62 | ICK | 2.20 | -1.10 |
| 20 | TEC | 1.61 | 0.29 | Pim2 | 2.19 | -0.27 |

**Kinase Score Plot (PTK left, STK right) of T1-96 vs T2-97**

