OXFORD

Systems biology

The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics

Danica D. Wiredja[1], Mehmet Koyutürk[2] and Mark R. Chance[1],*

[1]Department of Nutrition, Center for Proteomics and Bioinformatics and [2]Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106, USA

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Abstract

Summary: Computational characterization of differential kinase activity from phosphoproteomics datasets is critical for correctly inferring cellular circuitry and how signaling cascades are altered in drug treatment and/or disease. Kinase-Substrate Enrichment Analysis (KSEA) offers a powerful approach to estimating changes in a kinase's activity based on the collective phosphorylation changes of its identified substrates. However, KSEA has been limited to programmers who are able to implement the algorithms. Thus, to make it accessible to the larger scientific community, we present a web-based application of this method: the KSEA App. Overall, we expect that this tool will offer a quick and user-friendly way of generating kinase activity estimates from high-throughput phosphoproteomics datasets.

Availability and implementation: the KSEA App is a free online tool: casecpb.shinyapps.io/ksea/. The source code is on GitHub: github.com/casecpb/KSEA/. The application is also available as the R package 'KSEAapp' on CRAN: CRAN.R-project.org/package=KSEAapp/.

Contact: mark.chance@case.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

An understanding of kinase regulation is crucial for understanding many different biological signaling processes, the molecular pathogenesis of many diseases, and their potential reversals by kinase-altering therapies. Mass spectrometry-based phosphoproteomics has emerged as the leading high-throughput platform for measuring the identities and intensities of thousands of phosphopeptides simultaneously (Olsen and Mann, 2013). Consequently, there is a growing interest in generating bioinformatic tools to distill these highly complex datasets into biologically meaningful inferences of kinase activity changes. Currently available applications that offer such analyses include IKAP (Mischnik *et al.*, 2016), KinasePA (Yang *et al.*, 2016), CLUE (Yang *et al.*, 2015) and KEA (Lachmann and Ma'ayan, 2009), now updated as KEA2. However, the present implementation of IKAP is platform-specific, KinasePA and CLUE are limited to

multi-condition studies, and KEA is focused on substrate overrepresentation rather than kinase scoring.

As an alternative, Kinase–Substrate Enrichment Analysis (KSEA) scores each kinase based on the relative hyper-phosphorylation or dephosphorylation of the majority of its substrates, as identified from phosphosite-specific Kinase–Substrate (K–S) databases. The negative or positive value of the score, in turn, implies a decrease or increase in the kinase's overall activity relative to the control. Unfortunately, while KSEA offers a concise and easily interpretable scoring system, its accessibility remains restricted to programming experts, as the original source code was never released. Thus, this method has not been widely implemented due to the lack of a user-friendly tool. To make KSEA available to the greater scientific community, we present a web-based implementation: the KSEA App. This online tool is designed for users with wide-ranging

backgrounds who wish to identify and visualize kinase-level annotations from their quantitative phosphoproteomics datasets. We hope that this application would allow KSEA to become a routine, if not standard, analysis approach for phosphoproteomics.

2 Materials and methods

2.1 KSEA algorithm overview

The KSEA formula was previously described (Casado *et al.*, 2013). Assume that we are given a phosphoproteomics dataset with test and control samples, in which the fold change (FC) between test and control is computed for each phosphosite. As defined previously, the kinase's normalized score is calculated as follows:

$$score = \frac{(\bar{s} - \bar{p})\sqrt{m}}{\delta}$$

Here, $\bar{s}$ denotes the mean log2(FC) of known phosphosite substrates of the given kinase, $\bar{p}$ represents the mean log2(FC) of all phosphosites in the dataset, m denotes the total number of phosphosite substrates identified from the experiment that annotate to the specified kinase, and δ denotes the standard deviation of the log2(FC) across all phosphosites in the dataset. This formula is based on a z-score transformation, and we assume that the resulting scores (denoted as 'z-score' in the KSEA App outputs) are normally distributed. Subsequently, the *P*-value is determined by assessing the one-tailed probability of having a more extreme score than the one measured, followed by a Benjamini-Hochberg FDR correction for multiple hypothesis testing.

Interpretation of Results: The score of a kinase is based exclusively on the collective phosphorylation status of its substrates. Sites on the kinase itself are disregarded. For a FC = test/control, a kinase with a negative score has substrates that are generally dephosphorylated with the test group. This kinase, in turn, has decreased activity output in the test condition and is deemed downregulated. The inverse is true for positive scores.

2.2 Kinase–Substrate (K–S) dataset

To identify the substrates for each kinase, the KSEA App sources K–S annotations from PhosphoSitePlus (PSP) (Hornbeck *et al.*, 2012) and from NetworKIN (Linding *et al.*, 2007). PSP annotations are curated, and our current implementation uses annotations that are restricted to human proteins from the July 2016 release. In contrast, NetworKIN offers predicted relationships, which we downloaded as pre-computed data (calculated against human ENSEMBL version 59) from the KinomeXplorer-DB website (Horn *et al.*, 2014). By default, the KSEA App utilizes the PSP resource alone. However, users have the option to include NetworKIN annotations, and they can adjust the threshold on the NetworKIN confidence score to change the inclusion stringency. Additional explanations on the NetworKIN contributions and the KSEA formula are found in the Supplementary file 'Details on Methods.pdf'. Since the KSEA App utilizes downloaded K–S sources, we will annually release newer versions as the databases get updated.

2.3 Implementation

This KSEA App version 1.0 is hosted on the shinyapps.io server as a free online tool: https://casecpb.shinyapps.io/ksea/. The source code and local access details are found in https://github.com/casecpb/KSEA/. The **User Manual**, found within both sites, offers comprehensive instructions. Alternatively, this tool is available as the R package 'KSEAapp' in

CRAN: https://CRAN.R-project.org/package=KSEAapp/, and the source code details are found in https://github.com/casecpb/KSEApp/.

3 Results

3.1 Performance evaluation: KSEA without NetworKIN

KSEA was applied to a published quantitative phosphoproteomics dataset that studied MEK inhibition in lung adenocarcinoma (Kim *et al.*, 2016). We restricted the analysis to the A549 cell line differentially treated with selumetinib (AZD-6244), a highly selective MEK1/2 noncompetitive inhibitor, vs. DMSO control. As a start, K–S relationships were based solely on annotations from the PSP database. Fold changes were calculated by the ratio of selumetinib/ DMSO so that a negative kinase z-score represents relative collective dephosphorylation of substrates with selumetinib.

Following KSEA analysis, MEK1 (gene name: *MAP2K1*) exhibited statistically significant downregulation, along with some decrease in ERK1 (*MAPK3*), ERK2 (*MAPK1*) and RSK1 (*RPS6KA1*) signaling, as reflected in the negative z-scores (Supplementary Fig. S1, Table S1A). These results are consistent with (i) published enzymatic assays that demonstrated decreased MEK activity with selumetinib (Yeh *et al.*, 2007) and (ii) predicted blunting of effector kinases downstream of MEK, based on the canonical MAPK signaling pathway (Anjum and Blenis, 2008; Roberts and Der, 2007). These findings suggest that the KSEA App correctly identifies key kinase perturbations from this phosphoproteomics dataset.

3.2 Performance evaluation: KSEA + NetworKIN

Since only a small fraction of experimentally identified phosphosites have documented K–S annotations in PSP, the majority of the phosphoproteomics dataset could not be used in the previous KSEA calculations. Thus, to maximize the number of the phosphosites from the input that can be used in the calculations, we expanded the K–S database to include relationships predicted by NetworKIN. We then reran KSEA against this supplemented database with a NetworKIN score minimum of 5. This adjustment nearly doubled the total usable phosphosite count from 676 to 1327, which resulted in an improved 18% coverage. Furthermore, the number of scored kinases increased from 175 to 235. Permutation tests on the NetworKIN annotations are highlighted in Supplementary Figure S3.

Even with a large influx of new K–S predictions, many of the KSEA + NetworKIN results remained consistent with the previous findings. The MAPK signaling nodes that were downregulated in the earlier analysis retained the same directionality (Supplementary Fig. S2, Table S1C). More interestingly, however, EGFR showed statistically significant increase in activity output with the NetworKIN addition (Supplementary Fig. S2, Table S1C), whereas it did not meet the *P*-value cutoff before (Table S1A). This is due to the recruitment of 4 predicted EGFR substrates that exhibited strong hyperphosphorylation with drug (Table S1D). This finding, along with upregulation of *PDPK1* protein, is consistent with previous observations that noted enhanced phosphorylation through the *EGFR-PDPK1-AKT* axis (Kim *et al.*, 2016). Overall, based on our case study, the NetworKIN predictions improve phosphosite coverage and may boost the scores of kinases with few curated substrates in PSP.

Funding

Conflict of Interest: none declared.

References

Anjum,R. and Blenis,J. (2008) The RSK family of kinases: emerging roles in cellular signalling. *Nat. Rev. Mol. Cell Biol.*, **9**, 747–758.

Casado,P. *et al.* (2013) Kinase–substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal.*, **6**, rs6–rs6.

Horn,H. *et al.* (2014) KinomeXplorer: an integrated platform for kinome biology studies. *Nat. Methods*, **11**, 603–604.

Hornbeck,P.V. *et al.* (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.*, **40**, D261–D270.

Kim,J.-Y. *et al.* (2016) Phosphoproteomics Reveals MAPK Inhibitors Enhance MET- and EGFR-Driven AKT Signaling in KRAS-Mutant Lung Cancer. *Mol. Cancer Res.*, **14**, 1019–1029.

Lachmann,A. and Ma'ayan,A. (2009) KEA: Kinase enrichment analysis. *Bioinformatics*, **25**, 684–686.

Linding,R. *et al.* (2007) Systematic discovery of in vivo phosphorylation networks. *Cell*, **129**, 1415–1426.

Mischnik,M. *et al.* (2016) IKAP: A heuristic framework for inference of kinase activities from Phosphoproteomics data. *Bioinformatics*, **32**, 424–431.

Olsen,J.V. and Mann,M. (2013) Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol. Cell. Proteomics*, **12**, 3444–3452.

Roberts,P.J. and Der,C.J. (2007) Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*, **26**, 3291–3310.

Yang,P. *et al.* (2016) KinasePA: Phosphoproteomics data annotation using hypothesis driven kinase perturbation analysis. *Proteomics*, **16**, 1868–1871.

Yang,P. *et al.* (2015) Knowledge-Based Analysis for Detecting Key Signaling Events from Time-Series Phosphoproteomics Data. *PLoS Comput. Biol*, **11**, e1004403.

Yeh,T.C. *et al.* (2007) Biological Characterization of ARRY-142886 (AZD6244), a Potent, Highly Selective Mitogen-Activated Protein Kinase Kinase 1/2 Inhibitor. *Clin. Cancer Res.*, **13**, 1576–1583.