

In Silico Tools and Phosphoproteomic Software Exclusives

Authors:

Piby Paul, Manikandan Muthu, Yojitha Chilukuri, Steve W. Haga, Sechul Chun, Jae-Wook Oh

Date Submitted: 2020-01-02

Keywords: databases, tools, bioinformatics, data analysis, phosphoproteomics, proteomics

Abstract:

Proteomics and phosphoproteomics have been emerging as new dimensions of omics. Phosphorylation has a profound impact on the biological functions and applications of proteins. It influences everything from intrinsic activity and extrinsic executions to cellular localization. This post-translational modification has been subjected to detailed study and has been an object of analytical curiosity with the advent of faster instrumentation. The major strength of phosphoproteomic research lies in the fact that it gives an overall picture of the workforce of the cell. Phosphoproteomics gives deeper insights into understanding the mechanism behind development and progression of a disease. This review for the first time consolidates the list of existing bioinformatics tools developed for phosphoproteomics. The gap between development of bioinformatics tools and their implementation in clinical research is highlighted. The challenge facing progress is ideally believed to be the interdisciplinary arena this field of research is associated with. For meaningful solutions and deliverables, these tools need to be implemented in clinical studies for obtaining answers to pharmacodynamic questions, saving time, costs and energy. This review hopes to invoke some thought in this direction.

Record Type: Published Article

Submitted To: LAPSE (Living Archive for Process Systems Engineering)

Citation (overall record, always the latest version):

LAPSE:2020.0008

Citation (this specific file, latest version):

LAPSE:2020.0008-1

Citation (this specific file, this version):

LAPSE:2020.0008-1v1

DOI of Published Version: <https://doi.org/10.3390/pr7120869>

License: Creative Commons Attribution 4.0 International (CC BY 4.0)

Review

In Silico Tools and Phosphoproteomic Software Exclusives

Piby Paul ¹, Manikandan Muthu ², Yojitha Chilukuri ¹, Steve W. Haga ³, Sechul Chun ²  and Jae-Wook Oh ^{4,*} 

¹ St. Jude Childrens Cancer Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA; piby.paul@stjude.org (P.P.); yojitha.chilukuri@stjude.org (Y.C.)

² Department of Environmental Health Sciences, Konkuk University, Seoul 143-701, Korea; bhagatmani@gmail.com (M.M.); scchun@konkuk.ac.kr (S.C.)

³ Department of Computer Science and Engineering, National Sun Yat Sen University, Kaohsiung 804, Taiwan; stevewhaga@yahoo.com

⁴ Department of Stem Cell and Regenerative Biology, Konkuk University, Seoul 05029, Korea

* Correspondence: ohjw@konkuk.ac.kr; Tel.: +82-2-2049-6271; Fax: +82-2-455-1044

Received: 17 October 2019; Accepted: 18 November 2019; Published: 21 November 2019



Abstract: Proteomics and phosphoproteomics have been emerging as new dimensions of omics. Phosphorylation has a profound impact on the biological functions and applications of proteins. It influences everything from intrinsic activity and extrinsic executions to cellular localization. This post-translational modification has been subjected to detailed study and has been an object of analytical curiosity with the advent of faster instrumentation. The major strength of phosphoproteomic research lies in the fact that it gives an overall picture of the workforce of the cell. Phosphoproteomics gives deeper insights into understanding the mechanism behind development and progression of a disease. This review for the first time consolidates the list of existing bioinformatics tools developed for phosphoproteomics. The gap between development of bioinformatics tools and their implementation in clinical research is highlighted. The challenge facing progress is ideally believed to be the interdisciplinary arena this field of research is associated with. For meaningful solutions and deliverables, these tools need to be implemented in clinical studies for obtaining answers to pharmacodynamic questions, saving time, costs and energy. This review hopes to invoke some thought in this direction.

Keywords: proteomics; phosphoproteomics; data analysis; bioinformatics; databases; tools

1. Introduction

The recent few decades have seen an escalation in applying computer-based knowledge into life science applications, especially medicine. Currently, proteomics and bioinformatics are deeply rooted in biological sciences, so much so that it is hard to progress without this integration. Both these interdisciplinary approaches draw motivation from cross disciplines such as physics, chemistry, biology, computer science and engineering. Proteomics and bioinformatics have realized their full potential in various areas of biological sciences, especially when it comes to medicine. This interdisciplinary research has had an unequivocal impact on both fields, and through both fields has impacted the fundamental understanding and unravelling of core biological process affecting human health and welfare.

Bioinformatics uses computational approaches to answer theoretical and experimental queries in life sciences. The growth of the biotechnology industry has enormously impacted disease characterization, pharmaceutical discovery, clinical healthcare, forensics, molecular understanding and agriculture. These are core issues that fundamentally impact economic and social issues worldwide [1].

Incorporation of computer knowledge into biotechnological research has been responsible for taking things forward rapidly and authoritatively in this field. Scientific research has been on the transition in recent years owing to the collective information obtained from numerous genome projects and application of high-throughput technologies and mass spectrometry. The development of computational tools has not only roused hope, but has also provided increasing opportunities on biological systems [2]. Bioinformatics is now more so an empowering technology. A fundamental understanding on protein–protein interactions as well as protein identification and characterization and post translational modification has been achieved through bioinformatics approaches. The prediction of primary, secondary, tertiary and quaternary structures, and molecular modeling and visualization, has been realized through inputs from bioinformatics. Insights into genomics, epigenomics, lipidomics, glycomics, foodomics and transcriptomics has been the working hub of ongoing bioinformatics.

Phosphorylation is the chemical addition of a phosphoryl group (PO_3^-) to an organic molecule. The removal of a phosphoryl group is called dephosphorylation. Phosphorylation and dephosphorylation are carried out by kinases and phosphotransferases. Protein phosphorylation is the addition of a phosphoryl group to an amino acid. The amino acid is ideally serine, however, threonine and tyrosine in eukaryotes and histidine in prokaryotes are also on the list. The most predominant types of phosphorylation are post-translational modifications (PTM). The identification and characterization of proteins possessing phosphorylation as a post-translational modification (PTM) is phosphoproteomics. This branch of omics provides insights into proteins that regulate essential signaling pathways. It also aids in the understanding of cellular processes enabling the location of potential drug targets. Developments in sample preparation, enrichment, quantification and data analysis strategies have led to targeted phosphoproteome profiling. Using shotgun phosphoproteomics, enzymatic digestion of protein samples into peptides and phosphopeptides has been achieved. Phosphoproteomics has enabled the identification of site-specific phosphorylation in plants [3]. Technological advancements in analytical instrumentation, sample preparation and data analysis [4–14] have enabled obtaining high-quality, reproducible and comprehensive data sets. Researchers [15] were able to detect 50,000 phosphopeptides in a single human cancer cell line and quantify thousands of peptides within short time frames. Published reviews have elaborately discussed proteomics and phosphoproteomics in the context of precision medicine [16–18]. The relevance of phosphoproteomics data in providing mechanistic information towards the understanding of disease mechanism has been a crucial breakthrough [19–21]. The fundamental knowledge on the resistance of melanoma cells to BRAF inhibitors [19] as well as glioblastoma cells to mTOR inhibitors has been understood via phosphoproteomic studies. This profoundly insightful information has led to the discovery of novel combinational therapies [20]. Other authors [22] used phosphoproteomics data to assign tumor types for designing treatment routines. These same authors have studied acute myeloid leukemia primary cells to identify the differences in activation of kinases in cells and their drug resistance profiles [23]. Phosphoproteomics has led to unravelling the bidirectional signaling between endothelial cells and tumor cells for a better understanding of metastatic mechanisms of tumor cells [21]. Phosphoproteomic data has been employed to create mechanistic models of colorectal cancer cell lines for the understanding of specific drug resistance [24]. It is well-known that technological advancements and community efforts to standardize protocols and achieve reproducible results are vital for disease and patient stratification. Other than the data reproducibility issue that the mass spectrometry community is confronting, data type-specific methods to extract valuable information is another issue. The role of bioinformatics in proteomics/phosphoproteomics is thus evident: storage of huge volumes of information, cross examination and cross verification of patient sample information, simulation studies, simplification of in vivo/in vitro processes through theoretical approach and understanding underlying fundamental interactions and networking within diseased cells. Figure 1 gives the overall workflow of phosphoproteomics, indicating the junctures (data acquisition and data analysis) where bioinformatic tools play a pivotal role.

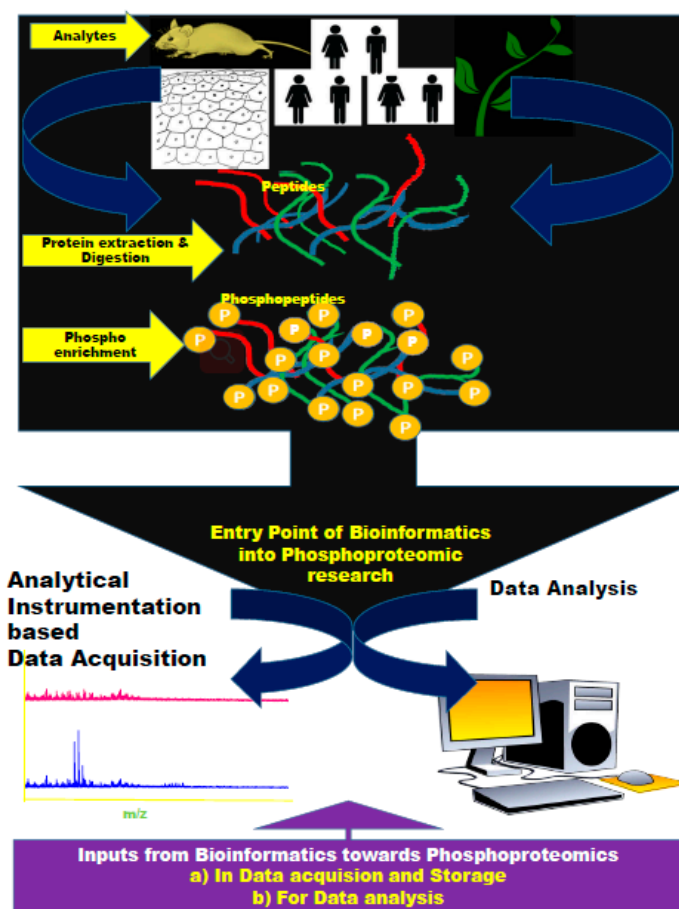


Figure 1. Work flow of phosphoproteomics indicating entry points where bioinformatics tools come handy.

The present review focusses on highlighting the importance of phosphoproteomic research and the importance of bioinformatics approaches and inputs into this area of research. The milestones achieved thus far via such an integration are presented and the challenges facing this integration discussed. This review discloses the fact that in spite of the valuable deliverables from phosphoproteomics, the interest from the research community in this area of omics is limited. Less than few tens of publications are placed on record; the need for implementation and the reason for this reduced popularity are also discussed in this review.

2. Biocomputational Tools for Proteomics—A Snapshot

With the increasingly large variety of proteomics workflows and data outcomes, Human Proteome Organisation (HUPO) [25] is facing a major challenge. It is here that there is room for a new generation of the ProteinScape™ bioinformatics platform, supported by LOOPP and PROCHECK software, to chip in. This could prove helpful in furthering functional characterization of specified proteins [26]. Other basic databases for proteins and genes include: UniProt Knowledgebase, Entrez Gene, OMIM Online Mendelian Inheritance in Man and Gene Ontology. Protein Interaction Databases include: DIP (Database of Interacting Proteins), BIND (Biomolecular Interaction Database Molecular interactions), IntAct, MIPS (Munich Information Center for Protein Sequences), HPRD (Human Protein Reference Database), STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) and MINT (Molecular Interaction). BioGRID, PIPs, MPIDB and TAIR, and additional tools such as PANTHER, DAVID, KEGG, and IPA, have been improved for data mapping. These tools are useful in understanding the functions of proteins in cells and its intricate interactions. Coon OMSSA Proteomic Analysis Software Suite (COMPASS) is a software that is freely available for high-throughput analysis of proteomics data,

based on the Open Mass Spectrometry Search Algorithm [27]. SPIRE (Systematic Protein Investigative Research Environment) has a web-interface that is easy to use, generating interactive and simple data formats [28] for mass spectrometric (MS) data. ScanRanker identifies unassigned high-quality spectra (that evaded identification) and picks spectra for de novo sequencing and cross-linking of proteins [29]. Also available are computer-based tools for biological pathways such as: iPath, Protein Lounge, BioCarta, KEGG and MetsCyc. Other software available for network analysis includes Ingenuity Pathway Analysis, MetaCore Integrated software suite based on MetaBase, PathwayStudio, GenMAPP (Gene Map Annotator and Pathway Profiler) and Cytoscape [2].

To transform large-scale biologically relevant proteomic data into valuable information [30], novel and improved computational tools are required. Various bioinformatics tools tailored to address the pressing needs of proteomics are available: Proteo Connections, Pathway Browser, and interaction databases like IntAct, ChEMBL, BioGRID [31] and ProteoRed MIAPE [32]. Proteomic storekeeper repositories like PRIDE, Global Proteome Machine, PeptideAtlas are also available to cater to huge volumes of mass spectral data and their respective protein identifications [33]. ANTILOPE is used for mathematical programming [34] and Peptidomimetics Based Inhibitor Design is a drug designing tool [35]. Genome Medicine Database of Japan Proteomics (GeMDBJ proteomics) is a free database [36] and HUPO [37] is another proteomic database that exchanges and imports data to and from databases such as Primer3 software [38], ClustalW, [39] SWISS 2DPAGE and others [40,41]. The MAPU (Max–Planck Unified Proteome Database) 2.0 database contains a huge collection of proteomes of organelles, tissues and cell types [42]. It aids in the retrieval of organism-specific proteomic data obtained from high accuracy MS-based proteomics and provides insight into general features ranging from gene ontology classification to SwissProt annotation. MODELLER 9v2 software is used to predict the 3-dimensional structure of proteins and PROCHECK and VERIFY 3D for generating output models [43,44].

Emerging tools used in various R&D sectors are summarized [45] as follows: (i) FindMod: Predicts post-translational modifications and single amino acid substitutions in peptides; (ii) FindPept: Identifies peptides; (iii) Mascot: Useful in protein identification by peptide mass fingerprinting; (iv) PepMAPPER: A web-based mapping tool developed for the purpose of epitope prediction and for sequence-structure alignment of proteins; (v) ProFound: Searches known protein sequences; (vi) ProteinProspector: Tools for peptide masses data (MS-Fit, MS-Pattern, MS-Digest); (vii) AACompIdent: Identifies a protein by its amino acid composition; (viii) AACompSim: Compares amino acid composition; (ix) TagIdent: Identifies proteins based on isoelectric point (pI), molecular weight (Mw) and sequence tag; (x) MultiIdent: Identifies proteins; (xi) InterPro Scan: Searches associated proteins with PROSITE, Pfam, PRINTS and other family and domain databases; (xii) MyHits: Establishes connectivity between protein sequences and motifs; (xiii) ScanProsite and HamapScan: Scans a sequence against PROSITE/HAMAP families; and finally, (xiv) MotifScan: Scans a sequence against protein profile databases [46].

The list further extends to tools [46] such as Pfam HMM search, ProDom, SUPERFAMILY Sequence Search, FingerPRINTScan, ELM (Eukaryotic Linear Motif) resource, PRATT, ChloroP, LipoP, MITOPROT, PATS (Prediction of apicoplast targeted sequences), PlasMit, Predotar, PTS1, SignalP, DictyOGlyc, NetCGlyc, ProtParam, Compute pI/Mw, ScanSite pI/Mw, MW, pI, Titration curve, HeliQuest, Radar, REP, REPRO, Homology modeling SWISS-MODEL, CPHmodels, ESyPred3D, Geno3d, Phyre (Successor of 3D-PSSM), Fugue, HHpred, SAM-T08, PSIPred, MakeMultimer, EBI PISA, PQS (Protein Quaternary Structure), ProtBud, Swiss-PdbViewer, SwissDock, EADock DSS and SwissParam. Tushar et al. [47] have extensively reviewed these tools in their review on the topic integration of bioinformatic tools for proteomics research.

3. Biocomputational Tools for Phosphoproteomics

Phosphoproteomic tools are being developed in such a way that involves searching a sequence database and performing analysis using designated tools [48,49]. Another approach employs searching a

spectral localization library [50–52]. We move forward to sweep through the existing phosphoproteomic software options available.

3.1. Tools for Analysis of Phosphopeptide Data/Spectra

SimPhospho, simulates phosphopeptide spectra searching through spectral libraries leading to highly accurate phosphosite validation. SimPhospho, accurately simulates phosphopeptide tandem mass spectra. The SimPhospho software uses Proteowizard project [53] and an XML library (Thomson) and includes a Qt framework-based user interface. Two XML files [54] serve as an input to SimPhospho: (i) a pep.xml file that contains search results [55] and (ii) an mzXML file that holds mass spectra. The software can be retrieved at <https://sourceforge.net/projects/simphospho/>.

The typical outcome from MS based proteomics, is the identification of peptides assigned to proteins. As a result of detection of extensive sub-proteomes and sub-phosphoproteomes of living cells, description, storage, management and recovery of the obtained data becomes challenging. For this purpose, PHOSIDA, the Phosphorylation Site Database [56] (<http://www.phosida.com>) was created [56]. The aim of PHOSIDA is to evolve high quality phosphoproteomic data for quantitative information, for mapping cell regulation after treatment with a stimulus. PHOSIDA is multifunctional in that it predicts putative phosphorylation sites, acetylation and other post-translational modification sites and analyzes phosphorylation events of proteins of interest. Computer based extraction of knowledge from comprehensive datasets is the agenda of ‘knowledge discovery in databases’ (KDD).

A large number of phosphopeptides and proteins are detected through mass spectrometry-based phosphoproteomics. The critical challenge is the manual analysis of downstream data. Towards this automation, a software called PhosFox [57] has been launched, which enables peptide-level processing of phosphoproteomic data supported by Mascot, Sequest, and Paragon. The PhosFox software aids in qualitative and quantitative phosphoproteomics studies and detects phosphorylated peptides and proteins. It also distinguishes differences within phosphorylation sites.

Normalization is a crucial step when analyzing phosphoproteomics data. A median normalization global centric method has been widely employed when it comes to label-free MS-based proteomics [58]. This works on the assumption that peptide abundances do not change between samples [59,60]. Researchers have reported that applying global-centering normalization introduces bias in distribution of fold changes of phosphopeptides across samples. It is in this direction that an R package called phosphonormalizer that fulfils pairwise normalization has been launched [61].

While thousands of phosphopeptides are identified in complex biological specimens, tools to evaluate and detect large amounts of phosphopeptides and related data are needed. Skyline is a freely-available and open source Windows client application for building Selected Reaction Monitoring (SRM)/Multiple Reaction Monitoring (MRM), Parallel Reaction Monitoring (PRM), Data Independent Acquisition (DIA/SWATH) and Data Dependent Acquisition (DDA) with MS1 quantitative methods and analyzing the resulting mass spectrometer data. MaxQuant [62] and Skyline [63] have been used in a few occasions for phosphopeptide identification and quantification. A couple of excellent reviews describe these software programs in more detail [64,65].

3.2. Tools for Phosphorylation Site Assignment

Correct phosphorylation site assignment is a critical aspect for any phosphoproteomic analysis. PhosphoScore [66] is such a site assignment program. It relates the match quality and intensity of observed spectral peaks compared to a theoretical spectrum. The claim [66] is that PhosphoScore produces >95% MS2 assignments. Ascore [67] is another statistical algorithm that measures the probability of correct phosphorylation sites. It is reported that phosphorylation sites with an Ascore ≥ 19 are usually considered unambiguously assigned.

3.3. Tools for Prediction of Phosphorylation Sites

As is known, protein phosphorylation is catalyzed by a group of enzymes called kinases, which add phosphate (PO_4) to serine (S), threonine (T), tyrosine (Y) and histidine (H) residues. On the other hand, phosphate moieties existing on substrates can also be eradicated by phosphatases. Since many members of the human protein kinase family are implicated in cancer, it is reported that their alteration or dysregulation provides clinically-validated targets for personalized treatment of cancer [68,69]. Given this fact, identification and characterization of kinases and their unique phosphorylation sites becomes a prerequisite for understanding protein kinase-regulated signaling pathways and their impacts on health and disease. While most or all protein kinases have been identified, the sites that they phosphorylate are not well understood. Many computational techniques for phosphorylation site prediction have been proposed. These differ in several ways, including the machine learning technique; the sequence information used; the number of residues surrounding the phosphorylation site; use of structural information/sequence information; and dependence on predictions made for specific/general kinases. Few review articles have previously been published that elaborately discuss computational phosphorylation site prediction. Kobe et al. [70] provided a brief review of this field [71], while Miller and Blom [72] briefly summarized the literature on phosphorylation site prediction and discussed their NetPhos [73,74] family of tools. Xue et al. [75] reviewed, and Trost and Kausalik [76] have extensively reviewed, the tools available for prediction of phosphorylation sites. The list includes tools such as: NetPhosK, PHOSITE, Predikin 1.0, DISPHOS, PredPhospho, GPS 1.0, GPS 2.1, KinasePhos 1.0, KinasePhos 2.0, NetworKIN, PhosPhAt, AutoMotif, PhoScan, Sitesseek, Predikin 2.0, Phos3D, PostMod, PPRED, Musite.

MusiteDeep [77], is an advanced deep-learning framework that predicts general and kinase-specific phosphorylation sites. DeepPhos [78], is another novel deep learning architecture for prediction of protein phosphorylation, applied for kinase-specific prediction. DeepPhos is reported to outperform competitive predictors in general and kinase-specific phosphorylation site prediction. PhosphoPredict [79] is yet another novel bioinformatics tool, which combines protein sequence and functional features to predict kinase-specific substrates and their associated sites.

3.4. Tools for Detection of Phosphosites and Kinase Activity from Phosphopeptide Data

Another approach to phosphoproteomics is through biochemical methods whereby kinase activities are assessed in vitro [80,81]. The major limitation is that these methods are limited in throughput and time-consuming. In vitro methods are not effective in reflecting in vivo activities of kinases, which is why MS-based methods are needed for evaluating kinase activity [82,83]. An approach to link phosphoproteomics data with the activity of kinases was presented by Qi et al. [84], which is known as kinase activity analysis (KAA). CLUE (CLUster Evaluation) is a method designed specifically for phosphoproteomics data [85], based on the hypothesis that phosphosites targeted by the same kinase will show similar temporal profiles. This principle has been utilized to guide the clustering algorithm and group kinases associated to these clusters. The abundances of the target phosphosites are studied using MS followed by in vitro enzymatic reactions. Since every phosphorylation event results from the activity of a kinase, the data thus involved is able to infer the activity of many kinases without the need of actual experiments. This task requires computational analysis of the detected phosphorylation sites (phosphosites), since thousands of phosphosites can routinely be measured in a single experiment. GSEA (Gene Set Enrichment Analysis), is generally applied to an entire set of gene expression data in order to obtain extensive information. It has also been reported to be useful for inference of kinase activity from phosphoproteomics data. This is related to the inference of transcription factor activity, based on the gene expression data.

There are many freely available databases that collect experimentally verified phosphosites, such as PhosphoSitePlus [86], Phospho.ELM [87], Signor [88], or PHOSIDA (explained above) [89]. Each of these databases differ in size and aim. For example, Phospho.ELM computes a score for the conservation of a phosphosite and Signor focuses on interactions with proteins involved in

signal transduction. PhosphoNetworks [90] is dedicated to kinase–substrate interactions. One other prominent database for interactions between kinases and individual phosphosites is PhosphoSitePlus. The unique database PhosphoGRID is exceptional in that it provides analogous information [91] for *Saccharomyces cerevisiae*. Specific information about phosphatase targets can be found in DEPOD [92]. As estimated, there are between 100,000 [93] and 500,000 possible phosphosites in the human proteome, and this has been the motivation for the development of computational tools to predict in vivo kinase–substrate relationships [94]. Scansite [95] uses position-specific scoring matrices (PSSMs) obtained by positional scanning of peptide libraries [96] or phage display methods [97]. Netphorest [98] classifies phosphorylation sites instead of predicting individual kinase–substrate links [75,98]. The software packages NetworKIN [99] (extended as KinomeXplorer [100]) and iGPS [101] combine information about kinase recognition motifs, in vivo phosphorylation sites and contextual information (STRING database [102–104]).

Currently available applications that offer kinase related analyses include inference of kinase activities from phosphoproteomics (IKAP) [105], kinase perturbation analysis (KinasePA) [85], CLUE [106] and Kinase Enrichment Analysis (KEA) [107], now updated as KEA2. IKAP is platform-specific, KinasePA and CLUE are limited to multi-condition studies and KEA is based on substrate overrepresentation. Kinase–Substrate Enrichment Analysis (KSEA) [108] scores each kinase based on the relative hyperphosphorylation or dephosphorylation of its substrates. To make KSEA available to the greater scientific community, a web-based implementation called the KSEA App has been developed. This KSEA App version 1.0 is hosted on the shinyapps.io server as a free online tool: <https://casecpb.shinyapps.io/ksea/>. Alternatively, this tool is also available as the R package ‘KSEAapp’ in CRAN: [https://CRAN.R-project.org/package=protect\\$relax\protect{\begingroup1\endgroup\@@over4}\\$KSEAapp/](https://CRAN.R-project.org/package=protect$relax\protect{\begingroup1\endgroup\@@over4}$KSEAapp/).

4. Future Direction—Implementation of Biocomputation Integrated Phosphoproteomics

As summarized above and in Table 1 [109–133], it is clearly evident that biocomputation had indeed played a vital role in the establishment of phosphoproteomics as a well accomplished offshoot of omics. However, it was observed that not much implementation of these tools towards applications has been reported. Most of the publications present the potential of the bioinformatics tools, or the development of these tools, and very few target the implementation of these for relevant applications. Few publications on applying phosphoproteomics for precision medicine have been reported. Additionally, not much progress has been made in applying any of these bioinformatics tools for phosphoproteomics in clinical or plant/animal biotechnological research.

We assume that the challenge could be owing to the fact that compared to proteomics, phosphoproteomics is a more specialized field requiring more expertise. This could be a reason for the inhibition of extensive research interest in this direction. Moreover, this field being a highly interdisciplinary field (with acquaintance in cross disciplinary fields such as molecular biology, computation, protein chemistry and informatics) this very aspect could be a limiting factor. However, with good progress in the development of such valuable bioinformatics tools being achieved, it is now high time that these resources are put to productive and real time applications and ultimate utility realized. This review points out to this lacuna that in spite of so many tools being developed, nothing much has been accomplished in terms of fundamental understanding of human diseases or animal/plant pathogenicity.

Except for a few reports on cancer related studies where bioinformatics phosphoproteomic approaches came handy, there appears to be no implementation. An interdisciplinary approach with cross disciplinary researchers collaborating will lead to positive progress and practical implication for harnessing wholesome benefits.

Table 1. Overview of software options available for phosphoproteomics.

Software Function/Application	Bioinformatics Tool	Specified Function	Website	Ref.
Analysis of phosphopeptide data/spectra	SimPhospho	Search, simulate phosphopeptide spectra and tandem mass spectra	https://sourceforge.net/projects/simphospho/	[53–55]
	PHOSIDA	Storage, management and recovery of phosphopeptide data, predicting putative phosphorylation sites, acetylation and other post-translational modification sites and analyses phosphorylation events of proteins of interest	http://www.phosida.com	[56,89]
	Prophossi	Automating expert validation of phosphopeptide–spectrum matches from tandem mass spectrometry	http://www.compbio.dundee.ac.uk/prophossi	[109]
	PhosFox	Peptide-level processing of phosphoproteomic data generated by Mascot, Sequest, and Paragon, qualitative and quantitative phosphoproteomics	https://bitbucket.org/phintsan/phosfox	[57]
	R package, Phospho normalizer	Normalization of phosphoproteomics data	https://bioconductor.org/packages/phosphonormalizer	[61]
Correct phosphorylation site assignment	PhosphoScore	Phosphorylation site assignment	https://omictools.com/phosphoscore-tool	[66]
	Ascore	Phosphorylation site assignment	http://ascore.med.harvard.edu/ascore.php	[67]
Phosphorylation site prediction	NetPhos	Machine learning methods, artificial neural networks (ANNs)	cbs.dtu.dk/services/NetPhos	[73,74]
	Scansite	Machine learning methods, position-specific scoring matrices (PSSMs) used	scansite.mit.edu	[95]
	Predikin 1.0	Structural analysis (SA) used	predikin.biosci.uq.edu.au	[110]
	DISPHOS	Logistic regression (LA) used	www.dabi.temple.edu/disphos	[111]
	NetPhosK	ANN used	cbs.dtu.dk/services/NetPhos	[74]
	PredPhospho	Support vector machines (SVMs) used	(website no longer accessible)	[112]
	PHOSITE	PSSM	(website no longer accessible)	[113]
	GPS 1.0	PSSM, Markov clustering (MC) used	gps.biocuckoo.org	[114]
	KinasePhos 1.0	Hidden Markov Model (HMM) used	kinasephos.mbc.nctu.edu.tw	[115]

Table 1. Cont.

Software Function/Application	Bioinformatics Tool	Specified Function	Website	Ref.
	PPSP	Bayesian probability (BP) based	ppsp.biocuckoo.org	[116]
	NetworKIN /KinomeXplorer	ANN, PSSM based	networkin.info	[100,101,117]
	KinasePhos 2.0	SVM	kinasephos2.mbc.nctu.edu.tw	[118]
	AutoMotif	SVM	(website no longer accessible)	[119]
	PhosPhAt	SVM	phosphat.mpimp-golm.mpg.de	[120]
	PhoScan	PSSM	bioinfo.au.tsinghua.edu.cn/phoscan	[121]
	MetaPredPS	Meta-predictor (MP)	metapred.biolead.org/MetaPredPS	[122]
	SiteSeek	Non specified	(no web implementation available)	[123]
	Predikin 2.0	HMM, SA	predikin.biosci.uq.edu.au	[124]
	GPS 2.0	PSSM, genetic algorithm (GA)	gps.biocuckoo.org	[125]
	CRPhos	Conditional random fields (CRF)	www.ptools.ua.ac.be/CRPhos	[126]
	Phos3D	SVM	phos3d.mpimp-golm.mpg.de	[127]
	PPRED	PSSM, SVM	ashiskb.info/research/ppred	[128]
	PAAS	PSSM	(website no longer accessible)	[129]
	PostMod	PSSM	pbil.kaist.ac.kr/PostMod	[130]
	GPS 2.1	PSSM, GA	gps.biocuckoo.org	[131]
	Musite	SVM	musite.sourceforge.net	[132]
	MusiteDeep	Predicting general and kinase-specific phosphorylation sites	https://github.com/duolinwang/MusiteDeep	[77]
	DeepPhos	Prediction of protein phosphorylation, kinase-specific prediction	https://github.com/USTCHILab/DeepPhos	[78]
	PhosphoPredict	Prediction of kinase-specific substrates and associated phosphorylation sites	http://phosphopredict.erc.monash.edu/	[79]

Table 1. Cont.

Software Function/Application	Bioinformatics Tool	Specified Function	Website	Ref.
Inference of kinase activity from phosphoproteomics data/detection of phosphosites	Kinase-Substrate Enrichment Analysis (KSEA)	Computational characterization of differential kinase activity from phosphoproteomics datasets	https://casecpb.shinyapps.io/ksea/	[108]
	CLUE (CLUster Evaluation) include IKAP, KinasePA, KAA (Kinase activity analysis) and KEA	Computational analysis of the detected phosphorylation sites (phosphosites)	https://omictools.com/clue-tool	[83,84,106–108]
	GSEA (Gene Set Enrichment Analysis)	Inference of kinase activity from phosphoproteomics data	http://software.broadinstitute.org/gsea/	[133]
	PhosphoSitePlus	Database for expert-edited and curated interactions between kinases and individual phosphosites	https://www.phosphosite.org/homeAction.action	[86]
	Phospho.ELM	Computes a score for the conservation of a phosphosite	http://phospho.elm.eu.org	[87]
	Signor	Focuses on interactions with proteins involved in signal transduction	https://signor.uniroma2.it/	[88]
	Netphorest	Classifies phosphorylation sites	http://www.netphorest.info/	[98]
	PhosphoGRID	Related information for <i>Saccharomyces cerevisiae</i>	https://phosphogrid.org/	[91]
	DEPhO phosphorylation database DEPOD	Supports phosphatase–kinase substrate networks	http://www.koehn.embl.de/depod	[92]

5. Conclusions

This review aimed at consolidating the bioinformatic tools available, giving a snapshot of the ones useful for proteomics and touching on the tools available for phosphoproteomics. Despite such valuable tools having been developed, in terms of real time application into clinical/pathological research and investigations we are not even close to accomplished. It is about time bioinformatics tool developers loop in with biologists and implement their tools.

Author Contributions: Conceptualization, Writing-Original Draft Preparation, P.P. and M.M.; Writing-Review & Editing, Y.C. and S.W.H.; Supervision, Funding Acquisition, S.C. and J.-W.O.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rao, V.S.; Das, S.K.; Rao, V.J.; Gedela, S. Recent developments in life sciences research: Role of Bioinformatics. *Afr. J. Biotechnol.* **2008**, *7*, 495–503.
2. Ulloa, A.R.; Rodríguez, R. Bioinformatic tools for proteomic data analysis: An overview. *Biotechnol. Appl.* **2008**, *25*, 312–319.
3. Nakagami, H.; Sugiyama, N.; Mochida, K.; Daudi, A.; Yoshida, Y.; Toyoda, T.; Tomita, M.; Ishihama, Y.; Shirasu, K. Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant Physiol.* **2010**, *153*, 1161–1174. [[CrossRef](#)]
4. Nilsson, T.; Mann, M.; Aebersold, R.; Yates, J.R., 3rd; Bairoch, A.; Bergeron, J.J. Mass spectrometry in high-throughput proteomics: Ready for the big time. *Nat. Methods* **2010**, *7*, 681–685. [[CrossRef](#)]
5. Zhou, L.; Wang, K.; Li, Q.; Nice, E.C.; Zhang, H.; Huang, C. Clinical proteomics-driven precision medicine for targeted cancer therapy: Current overview and future perspectives. *Expert Rev. Proteom.* **2016**, *13*, 367–381. [[CrossRef](#)] [[PubMed](#)]
6. Guerin, M.; Gonçalves, A.; Toiron, Y.; Baudelet, E.; Audebert, S.; Boyer, J.B.; Borg, J.P.; Camoin, L. How may targeted proteomics complement genomic data in breast cancer? *Expert Rev. Proteom.* **2017**, *14*, 43–54. [[CrossRef](#)] [[PubMed](#)]
7. Mitchell, P. Proteomics retrenches. *Nat. Biotechnol.* **2010**, *28*, 665–670. [[CrossRef](#)]
8. Searle, B.C. Scaffold: A bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **2010**, *10*, 1265–1269. [[CrossRef](#)] [[PubMed](#)]
9. Varjosalo, M.; Sacco, R.; Stukalov, A.; van Drogen, A.; Planavsky, M.; Hauri, S.; Aebersold, R.; Bennett, K.L.; Colinge, J.; Gstaiger, M.; et al. Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. *Nat. Methods* **2013**, *10*, 307–314. [[CrossRef](#)] [[PubMed](#)]
10. Mann, M. Comparative analysis to guide quality improvements in proteomics. *Nat. Methods* **2009**, *6*, 717–719. [[CrossRef](#)] [[PubMed](#)]
11. Stead, D.A.; Paton, N.W.; Missier, P.; Embury, S.M.; Hedeler, C.; Jin, B.; Brown, A.J.; Preece, A. Information quality in proteomics. *Brief. Bioinform.* **2008**, *9*, 174–188. [[CrossRef](#)] [[PubMed](#)]
12. Tabb, D.L. Quality assessment for clinical proteomics. *Clin. Biochem.* **2013**, *46*, 411–420. [[CrossRef](#)] [[PubMed](#)]
13. Wang, X. Statistical assessment of QC metrics on raw LC-MS/MS data. Proteomics. In *Proteomics*; Comai, L., Katz, J.E., Mallick, P., Eds.; Springer: New York, NY, USA, 2017; pp. 325–337.
14. Whiteaker, J.R.; Halusa, G.N.; Hoofnagle, A.N.; Sharma, V.; MacLean, B.; Yan, P.; Wrobel, J.A.; Kennedy, J.; Mani, D.R.; Zimmerman, L.J.; et al. Using the CPTAC Assay Portal to identify and implement highly characterized targeted proteomics assays. *Methods Mol Biol.* **2016**, *1410*, 223–236.
15. Sharma, K.; D'Souza, R.C.J.; Tyanova, S.; Schaab, C.; Wiśniewski, J.R.; Cox, J.; Mann, M. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.* **2014**, *8*, 1583–1594. [[CrossRef](#)] [[PubMed](#)]
16. Casado, P.; Hijazi, M.; Britton, D.; Cutillas, P.R. Impact of phosphoproteomics in the translation of kinase-targeted therapies. *Proteomics* **2017**, *17*, 1600235. [[CrossRef](#)] [[PubMed](#)]
17. Cutillas, P.R. Role of phosphoproteomics in the development of personalized cancer therapies. *Proteom. Clin. Appl.* **2015**, *9*, 383–395. [[CrossRef](#)]

18. Yang, J.-Y.; Yoshihara, K.; Tanaka, K.; Hatae, M.; Masuzaki, H.; Itamochi, H.; Cancer Genome Atlas (TCGA) Research Network; Takano, M.; Ushijima, K.; Tanyi, J.L. Predicting time to ovarian carcinoma recurrence using protein markers. *J. Clin. Investig.* **2013**, *123*, 3740–3750. [[CrossRef](#)]
19. Parker, R.; Vella, L.J.; Xavier, D.; Amirkhani, A.; Parker, J.; Cebon, J.; Molloy, M.P. Phosphoproteomic analysis of cell-based resistance to BRAF inhibitor therapy in melanoma. *Front. Oncol.* **2015**, *5*, 95. [[CrossRef](#)]
20. Wei, W.; Shin, Y.S.; Xue, M.; Matsutani, T.; Masui, K.; Yang, H.; Ikegami, S.; Gu, Y.; Herrmann, K.; Johnson, D.; et al. Single-cell phosphoproteomics resolves adaptive signaling dynamics and informs targeted combination therapy in glioblastoma. *Cancer Cell* **2016**, *29*, 563–573. [[CrossRef](#)]
21. Locard-Paulet, M.; Lim, L.; Veluscek, G.; McMahon, K.; Sinclair, J.; van Weverwijk, A.; Worboys, J.D.; Yuan, Y.; Isacke, C.M.; Jørgensen, C. Phosphoproteomic analysis of interacting tumor and endothelial cells identifies regulatory mechanisms of transendothelial migration. *Sci. Signal.* **2016**, *9*, ra15. [[CrossRef](#)]
22. Casado, P.; Alcolea, M.P.; Iorio, F.; Rodríguez-Prados, J.C.; Vanhaesebroeck, B.; Saez-Rodriguez, J.; Joel, S.; Cutillas, P.R. Phosphoproteomics data classify hematological cancer cell lines according to tumor type and sensitivity to kinase inhibitors. *Genome Biol.* **2013**, *14*, R37. [[CrossRef](#)] [[PubMed](#)]
23. Casado, P.; Rodríguez-Prados, J.-C.; Cosulich, S.C.; Guichard, S.; Vanhaesebroeck, B.; Joel, S.; Cutillas, P.R. Kinase substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal.* **2013**, *6*, rs6. [[CrossRef](#)] [[PubMed](#)]
24. Eduati, F.; Doldan-Martelli, V.; Klinger, B.; Cokelaer, T.; Sieber, A.; Kogera, F.; Dorel, M.; Garnett, M.J.; Blüthgen, N.; Saez-Rodriguez, J. Drug resistance mechanisms in colorectal cancer dissected with cell type-specific dynamic logic models. *Cancer Res.* **2017**, *77*, 3364–3375. [[CrossRef](#)] [[PubMed](#)]
25. Thiele, H.; Jörg, G.; Peter, H.; Gerhard, K.; Martin, B. Managing Proteomics Data: From Generation and Data Warehousing to Central Data Repository. *J. Proteom. Bioinform.* **2008**, *1*, 485–507. [[CrossRef](#)]
26. Subramanian, R.; Muthurajan, R.; Ayyanar, M. Comparative Modeling and Analysis of 3-D Structure of EMV2, a Late Embryogenesis Abundant Protein of Vigna Radiata (Wilczek). *J. Proteom. Bioinform.* **2008**, *1*, 401–407.
27. Wenger, C.D.; Phanstiel, D.H.; Lee, M.V.; Bailey, D.J.; Coon, J.J. COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* **2011**, *11*, 1064–1074. [[CrossRef](#)]
28. Kolker, E.; Higdon, R.; Welch, D.; Bauman, A.; Stewart, E.; Haynes, W.; Broomall, W.; Kolker, N. SPIRE: Systematic protein investigative research environment. *J. Proteom.* **2011**, *75*, 122–126. [[CrossRef](#)]
29. Ma, Z.Q.; Chambers, M.C.; Ham, A.J.; Cheek, K.L.; Whitwell, C.W.; Aerni, H.R.; Schilling, B.; Miller, A.W.; Caprioli, R.M.; Tabb, D.L. ScanRanker: Quality assessment of tandem mass spectra via sequence tagging. *J. Proteom. Res.* **2011**, *10*, 2896–2904. [[CrossRef](#)]
30. Courcelles, M.; Lemieux, S.; Voisin, L.; Meloche, S.; Thibault, P. ProteoConnections: A bioinformatics platform to facilitate proteome and phosphoproteome analyses. *Proteomics* **2011**, *11*, 2654–2671. [[CrossRef](#)]
31. Haw, R.; Hermjakob, H.; D'Eustachio, P.; Stein, L. Reactome Pathway Analysis to Enrich Biological Discovery in Proteomics Datasets. *Proteomics* **2011**, *11*, 3598–3613. [[CrossRef](#)]
32. Medina-Aunon, J.A.; Martinez-Bartolome, S.; Lopez-Garcia, M.A.; Salazar, E.; Navajas, R.; Jones, A.; Paradela, A.; Albar, J. The Proteored MIAPE Web Toolkit: A User-Friendly Framework to Connect and Share Proteomics Standards. *Mol. Cell Proteom.* **2011**, *10*, 8334. [[CrossRef](#)]
33. Ponomarenko, E.A.; Ilgisonis, A.; Lisitsa, A.V. Knowledge-based technologies in proteomics. *Bioorg. Khim.* **2001**, *37*, 190–198. [[CrossRef](#)] [[PubMed](#)]
34. Sandro, A.; Gunnar, W.K.; Knut, R. Antilope—A Lagrangian Relaxation Approach to the de novo Peptide Sequencing Problem. *IEEE/ACM Trans Comput. Biol. Bioinform.* **2012**, *9*, 385–394.
35. Vetrivel, U.; Sankar, P.; Nagarajan, N.K.; Subramanian, G. Peptidomimetics Based Inhibitor Design for HIV-1 gp120 Attachment Protein. *J. Proteom. Bioinform.* **2009**, *2*, 481–484. [[CrossRef](#)]
36. Kikuta, K.; Tsunehiro, Y.; Yoshida, A.; Tochigi, N.; Hirohahsi, S.; Kawai, A.; Kondo, T. Proteome Expression Database of Ewing sarcoma: A segment of the Genome Medicine Database of Japan Proteomics. *J. Proteom. Bioinform.* **2009**, *2*, 500–504. [[CrossRef](#)]
37. Sandra, O.; Henning, H. Standardising Proteomics Data—the work of the HUPO Proteomics Standards Initiative. *J. Proteom. Bioinform.* **2008**, *1*, 3–5.
38. Neha, G.; Sachin, P.; Anil, P.; Anil, K. Primer Designing for Dreb1A, A Cold Induced Gene. *J. Proteom. Bioinform.* **2008**, *1*, 28–35.

39. Allam, A.R.; Kiran, K.R.; Hanuman, T. Bioinformatic Analysis of Alzheimer's Disease Using Functional Protein Sequences. *J. Proteom. Bioinform.* **2008**, *1*, 036–042. [[CrossRef](#)]
40. Kush, A.; Raghava, G.P.S. AC2DGel: Analysis and Comparison of 2D Gels. *J. Proteom. Bioinform.* **2008**, *1*, 43–46. [[CrossRef](#)]
41. Seenivasagan, R.; Kasimani, R.; Marimuthu, P.; Kalidoss, R.; Shanmughavel, P. Comparative Modeling of Viral Protein R (Vpr) From Human Immunodeficiency Virus 1 (Hiv 1). *J. Proteom. Bioinform.* **2008**, *1*, 73–76.
42. Gnad, F.; Oroshi, M.; Birney, E.; Mann, M. MAPU 2.0: High-accuracy proteomes mapped to genomes. *Nucleic Acids Res.* **2009**, *37*, D902–D906. [[CrossRef](#)] [[PubMed](#)]
43. Sunil, K.; Priya, R.D.; Prakash, C.S. Prediction of 3-Dimensional Structure of Cathepsin L Protein of Rattus Norvegicus. *J. Proteom. Bioinform.* **2008**, *1*, 307–314.
44. Paul, K.; Nathan, L.C.; Daniel, C.; Clarissa, D.; Patrice, H.; Joanna, H.; Eustache, P.; Marc, R.; Olivier, M.; Howard, M.C.; et al. Global Proteomics: Pharmacodynamic Decision Making via Geometric Interpretations of Proteomic Analyses. *J. Proteom. Bioinform.* **2008**, *1*, 315–328.
45. ExPASy. *SIB Bioinformatics Resource Portal-Proteomics Tools.html*.
46. Nanda, T.; Tripathy, K.; Ashwin, P. Integration of Bioinformatics Tools for Proteomics Research. *J. Comput. Sci. Syst. Biol.* **2001**, S13. [[CrossRef](#)]
47. Beausoleil, S.A.; Villén, J.; Gerber, S.A.; Rush, J.; Gygi, S.P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **2006**, *24*, 1285–1292. [[CrossRef](#)]
48. Olsen, J.V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127*, 635–648. [[CrossRef](#)]
49. Taus, T.; Köcher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* **2001**, *10*, 5354–5362. [[CrossRef](#)] [[PubMed](#)]
50. Bodenmiller, B.; Campbell, D.; Gerriets, B.; Lam, H.; Jovanovic, M.; Picotti, P.; Schlapbach, R.; Aebersold, R. PhosphoPep—A database of protein phosphorylation sites in model organisms. *Nat. Biotechnol.* **2008**, *26*, 1339–1340. [[CrossRef](#)]
51. Hummel, J.; Niemann, M.; Wienkoop, S.; Schulze, W.; Steinhäuser, D.; Selbig, J.; Walther, D.; Weckwerth, W. ProMEX: A mass spectral reference database for proteins and protein phosphorylation sites. *BMC Bioinform.* **2007**, *8*, 216. [[CrossRef](#)]
52. Suni, V.; Imanishi, S.Y.; Maiolica, A.; Aebersold, R.; Corthals, G.L. Confident site localization using a simulated phosphopeptide spectral library. *J. Proteome Res.* **2015**, *14*, 2348–2359. [[CrossRef](#)]
53. Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics* **2008**, *24*, 2534–2536. [[CrossRef](#)] [[PubMed](#)]
54. Keller, A.; Eng, J.; Zhang, N.; Li, X.J.; Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* **2005**, *1*, 2005.0017. [[CrossRef](#)] [[PubMed](#)]
55. Craig, R.; Beavis, R.C. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, 1466–1467. [[CrossRef](#)] [[PubMed](#)]
56. Florian, G.; Shubin, R.; Juergen, C.; Jesper, V.O.; Boris, M.; Mario, O.; Matthias, M. PHOSIDA (Phosphorylation Site Database): Management, Structural and Evolutionary Investigation, and Prediction of Phosphosites. *Genome Biol.* **2007**, *8*. [[CrossRef](#)]
57. Söderholm, S.; Hintsanen, P.; Öhman, T.; Aittokallio, T.; Nyman, T.A. PhosFox: A bioinformatics tool for peptide-level processing of LC-MS/MS-based phosphoproteomic data. *Proteome Sci.* **2014**, *12*, 36. [[CrossRef](#)]
58. Välikangas, T.; Suomi, T.; Elo, L.L. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinform.* **2016**, *19*, 1–11. [[CrossRef](#)]
59. Kauko, O.; Teemu, D.L.; Mikael, J.; Petteri, H.; Veronika, S.; Pekka, H.; Garry, C.; Tero, A.; Jukka, W.; Kauko, O. Label-free quantitative phosphoproteomics with novel pairwise abundance normalization reveals synergistic RAS and CIP2A signaling. *Sci. Rep.* **2015**, *5*, 13099. [[CrossRef](#)]
60. Olsen, J.V.; Mann, M. Status of large-scale analysis of posttranslational modifications by mass spectrometry. *Mol. Cell. Proteom.* **2013**, *12*, 3444–3452. [[CrossRef](#)]
61. Saraei, S.; Suomi, T.; Kauko, O.; Elo, L.L.; Stegle, O. Phosphonormalizer: An R package for normalization of MS-based label-free phosphoproteomics. *Bioinformatics* **2018**, *34*, 693–694. [[CrossRef](#)]
62. Tikira, T.; Juergen, C. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301–2319.

63. MacLean, B.; Tomazela, D.M.; Shulman, N.; Chambers, M.; Finney, G.L.; Frewen, B.; Kern, R.; Tabb, D.L.; Liebler, D.C.; MacCoss, M.J. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966–968. [[CrossRef](#)] [[PubMed](#)]
64. Tyanova, S.; Temu, T.; Carlson, A.; Sinitcyn, P.; Mann, M.; Cox, J. Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **2015**, *15*, 1453–1456. [[CrossRef](#)] [[PubMed](#)]
65. Pino, L.K.; Searle, B.C.; Bollinger, J.G.; Nunn, B.; MacLean, B.; MacCoss, M.J. The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom. Rev.* **2017**, 1–16. [[CrossRef](#)] [[PubMed](#)]
66. Ruttenberg, B.E.; Pisitkun, T.; Knepper, M.A.; Hoffert, J.D. PhosphoScore: An open-source phosphorylation site assignment tool for MSn data. *J. Proteome Res.* **2008**, *7*, 3054–3059. [[CrossRef](#)] [[PubMed](#)]
67. Hoffert, J.D.; Wang, G.; Pisitkun, T.; Shen, R.F.; Knepper, M.A. An automated platform for analysis of phosphoproteomic datasets: Application to kidney collecting duct phosphoproteins. *J. Proteome. Res.* **2007**, *9*, 3501–3508. [[CrossRef](#)]
68. Fleuren, E.D.; Zhang, L.; Wu, J.; Daly, R.J. The kinome ‘at large’ in cancer. *Nat. Rev. Cancer* **2016**, *16*, 83–98. [[CrossRef](#)]
69. Creixell, P.; Erwin, M.; Craig, D.S.; James, L.; Chad, J.M.; Hua, J.L.; Lara, P.; Thomas, R.C.; Nevena, Z.; Antonio, P.; et al. Kinome-wide decoding of network-attacking mutations rewiring cancer signaling. *Cell* **2015**, *163*, 202–217. [[CrossRef](#)]
70. Kobe, B.; Kamppmann, T.; Forwood, J.K.; Listwan, P.; Brinkworth, R.I. Substrate specificity of protein kinases and computational prediction of substrates. *Biochim. Biophys. Acta* **2005**, *1754*, 200–209. [[CrossRef](#)]
71. Hjerrild, M.; Gammeltoft, S. Phosphoproteomics toolbox: Computational biology, protein chemistry and mass spectrometry. *FEBS Lett.* **2006**, *580*, 4764–4770. [[CrossRef](#)]
72. Miller, M.L.; Blom, N. Kinase-specific prediction of protein phosphorylation sites. *Methods Mol. Biol.* **2009**, *527*, 299–310.
73. Blom, N.; Gammeltoft, S.; Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **1999**, *294*, 1351–1362. [[CrossRef](#)] [[PubMed](#)]
74. Hjerrild, M.; Stensballe, A.; Rasmussen, T.E.; Kofoed, C.B.; Blom, N.; Sicheritz-Ponten, T.; Larsen, M.R.; Brunak, S.; Jensen, O.N.; Gammeltoft, S. Identification of phosphorylation sites in protein kinase a substrates using artificial neural networks and mass spectrometry. *J. Proteome Res.* **2004**, *3*, 426–433. [[CrossRef](#)] [[PubMed](#)]
75. Xue, Y.; Gao, X.; Cao, J.; Liu, Z.; Jin, C.; Wen, L.; Yao, X.; Ren, J. A summary of computational resources for protein phosphorylation. *Curr. Protein Pept. Sci.* **2010**, *11*, 485–496. [[CrossRef](#)] [[PubMed](#)]
76. Trost, B.; Kusalik, A. Computational prediction of eukaryotic phosphorylation sites. *Bioinformatics* **2011**, *27*, 2927–2935. [[CrossRef](#)]
77. Wang, D.; Zeng, S.; Xu, C.; Qiu, W.; Liang, Y.; Joshi, T.; Xu, D. MusiteDeep: A deep-learning framework for general and kinase-specific phosphorylation site prediction. *Bioinformatics* **2017**, *33*, 3909–3916. [[CrossRef](#)]
78. Luo, F.; Wang, M.; Liu, Y.; Zhao, X.; Li, A. DeepPhos: Prediction of protein phosphorylation sites with deep learning. *Bioinformatics* **2019**, *35*, 2766–2773. [[CrossRef](#)]
79. Song, J.; Wang, H.; Wang, J.; Leier, A.; Marquez-Lago, T.; Yang, B.; Zhang, Z.; Akutsu, T.; Webb, G.; Daly, R.J. PhosphoPredict: A bioinformatics tool for prediction of human kinase-specific phosphorylation substrates and sites by integrating heterogeneous feature selection. *Sci. Rep.* **2017**, *7*, 6862. [[CrossRef](#)]
80. Newman, R.H.; Zhang, J.; Zhu, H. Toward a systems-level view of dynamic phosphorylation networks. *Front. Genet.* **2014**, *5*, 263. [[CrossRef](#)]
81. Glickman, J.F. *Assay Development for Protein Kinase Enzymes*; Eli Lilly & Company and the National Center for Advancing Translational Sciences: Bethesda, MD, USA, 2012.
82. Cutillas, P.R.; Khwaja, A.; Graupera, M.; Pearce, W.; Gharbi, S.; Waterfield, M.; Vanhaesebroeck, B. Ultrasensitive and absolute quantification of the phosphoinositide 3-kinase/Akt signal transduction pathway by mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8959–8964. [[CrossRef](#)]
83. Yu, Y.; Anjum, R.; Kubota, K.; Rush, J.; Villen, J.; Gygi, S.P. A site-specific, multiplexed kinase activity assay using stable-isotope dilution and high-resolution mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11606–11611. [[CrossRef](#)]

84. Qi, L.; Liu, Z.; Wang, J.; Cui, Y.; Guo, Y.; Zhou, T.; Zhou, Z.; Guo, X.; Xue, Y.; Sha, J. Systematic analysis of the phosphoproteome and kinase-substrate networks in the mouse testis. *Mol. Cell. Proteom.* **2014**, *13*, 3626–3638. [[CrossRef](#)] [[PubMed](#)]
85. Yang, P.; Zheng, X.; Jayaswal, V.; Hu, G.; Yang, J.Y.H.; Jothi, R. Knowledge Based Analysis for Detecting Key Signaling Events from Time-Series Phosphoproteomics Data. *PLoS Comput. Biol.* **2015**, *11*, e1004403. [[CrossRef](#)] [[PubMed](#)]
86. Hornbeck, P.V.; Zhang, B.; Murray, B.; Kornhauser, J.M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **2015**, *43*, D512–D520. [[CrossRef](#)] [[PubMed](#)]
87. Dinkel, H.; Chica, C.; Via, A.; Gould, C.M.; Jensen, L.J.; Gibson, T.J.; Diella, F. Phospho.ELM: A database of phosphorylation sites—update 2011. *Nucleic Acids Res.* **2011**, *39*, D261–D267. [[CrossRef](#)]
88. Perfetto, L.; Briganti, L.; Calderone, A.; Cerquone Perpetuini, A.; Iannuccelli, M.; Langone, F.; Licata, L.; Marinkovic, M.; Mattioni, A.; Pavlidou, T.; et al. SIGNOR: A database of causal relationships between biological entities. *Nucleic Acids Res.* **2016**, *44*, D548–D554. [[CrossRef](#)]
89. Gnad, F.; Gunawardena, J.; Mann, M. PHOSIDA 2011: The posttranslational modification database. *Nucleic Acids Res.* **2011**, *39*, D253–D260. [[CrossRef](#)]
90. Hu, J.; Rho, H.S.; Newman, R.H.; Zhang, J.; Zhu, H.; Qian, J. PhosphoNetworks: A database for human phosphorylation networks. *Bioinformatics* **2014**, *30*, 141–142. [[CrossRef](#)]
91. Sadowski, I.; Breitkreutz, B.J.; Stark, C.; Su, T.C.; Dahabieh, M.; Raithatha, S.; Bernhard, W.; Oughtred, R.; Dolinski, K.; Barreto, K.; et al. The PhosphoGRID *Saccharomyces cerevisiae* protein phosphorylation site database: Version 2.0 update. *Database (Oxford)* **2013**, *2013*, bat026. [[CrossRef](#)]
92. Duan, G.; Li, X.; Köhn, M. The human DEPhOsporylation database DEPOD: A 2015 update. *Nucleic Acids Res.* **2015**, *43*, D531–D535. [[CrossRef](#)]
93. Zhang, H.; Zha, X.; Tan, Y.; Hornbeck, P.V.; Mastrangelo, A.J.; Alessi, D.R.; Polakiewicz, R.D.; Comb, M.J. Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs. *J. Biol. Chem.* **2002**, *277*, 39379–39387. [[CrossRef](#)]
94. Lemeer, S.; Heck, A.J. The phosphoproteomics data explosion. *Curr. Opin. Chem. Biol.* **2009**, *13*, 414–420. [[CrossRef](#)] [[PubMed](#)]
95. Obenaus, J.C.; Cantley, L.C.; Yaffe, M.B. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* **2003**, *31*, 3635–3641. [[CrossRef](#)] [[PubMed](#)]
96. Chen, C.; Turk, B.E. Analysis of serine-threonine kinase specificity using arrayed positional scanning peptide libraries. *Curr. Protoc. Mol. Biol.* **2010**. [[CrossRef](#)] [[PubMed](#)]
97. Sidhu, S.S.; Koide, S. Phage display for engineering and analyzing protein interaction interfaces. *Curr. Opin. Struct. Biol.* **2007**, *17*, 481–487. [[CrossRef](#)]
98. Miller, M.L.; Jensen, L.J.; Diella, F.; Jørgensen, C.; Tinti, M.; Li, L.; Hsiung, M.; Parker, S.A.; Bordeaux, J.; Sicheritz-Ponten, T.; et al. Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.* **2008**, *1*, ra2. [[CrossRef](#)]
99. Linding, R.; Jensen, L.J.; Pasculescu, A.; Olhovsky, M.; Colwill, K.; Bork, P.; Yaffe, M.B.; Pawson, T. NetworKIN: A resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **2008**, *36*, D695–D699. [[CrossRef](#)]
100. Horn, H.; Schoof, E.M.; Kim, J.; Robin, X.; Miller, M.L.; Diella, F.; Palma, A.; Cesareni, G.; Jensen, L.J.; Linding, R. KinomeXplorer: An integrated platform for kinome biology studies. *Nat. Methods* **2014**, *1*, 603–604. [[CrossRef](#)]
101. Song, C.; Ye, M.; Liu, Z.; Cheng, H.; Jiang, X.; Han, G.; Songyang, Z.; Tan, Y.; Wang, H.; Ren, J.; et al. Systematic analysis of protein phosphorylation networks from phosphoproteomic data. *Mol. Cell. Proteom.* **2012**, *11*, 1070–1083. [[CrossRef](#)]
102. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K.P.; et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **2015**, *43*, D447–D452. [[CrossRef](#)]
103. Wagih, O.; Sugiyama, N.; Ishihama, Y.; Beltrao, P. Uncovering Phosphorylation-Based Specificities through Functional Interaction Networks. *Mol. Cell. Proteom.* **2016**, *15*, 236–245. [[CrossRef](#)]
104. Wirbel, J.; Cutillas, P.; Saez-Rodriguez, J. Phosphoproteomics-Based Profiling of Kinase Activities in Cancer Cells. *Methods Mol. Biol.* **2018**, *1711*, 103–132. [[CrossRef](#)] [[PubMed](#)]

105. Mischnik, M.; Sacco, F.; Cox, J.; Schneider, H.C.; Schäfer, M.; Hendlich, M.; Crowther, D.; Mann, M.; Klabunde, T. IKAP: A heuristic framework for inference of kinase activities from Phosphoproteomics data. *Bioinformatics* **2016**, *32*, 424–431. [[CrossRef](#)] [[PubMed](#)]
106. Yang, P.; Patrick, E.; Humphrey, S.J.; Ghazanfar, S.; James, D.E.; Jothi, R.; Yang, J.Y. KinasePA: Phosphoproteomics data annotation using hypothesis driven kinase perturbation analysis. *Proteomics* **2016**, *16*, 1868–1871. [[CrossRef](#)] [[PubMed](#)]
107. Lachmann, A.; Ma'ayan, A. KEA: Kinase enrichment analysis. *Bioinformatics* **2009**, *25*, 684–686. [[CrossRef](#)]
108. Wiredja, D.D.; Koyutürk, M.; Chance, M.R. The KSEA App: A web-based tool for kinase activity inference from quantitative phosphoproteomics. *Bioinformatics* **2017**, *33*, 3489–3491. [[CrossRef](#)]
109. Martin, D.M.; Nett, I.R.; Vandermoere, F.; Barber, J.D.; Morrice, N.A.; Ferguson, M.A. ProPhosSI: Automating expert validation of phosphopeptide-spectrum matches from tandem mass spectrometry. *Bioinformatics* **2010**, *26*, 2153–2159. [[CrossRef](#)]
110. Brinkworth, R.I.; Breinl, R.A.; Kobe, B. Structural basis and prediction of substrate specificity in protein serine/threonine kinases. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 74–79. [[CrossRef](#)]
111. Iakoucheva, L.M.; Radivojac, P.; Brown, C.J.; O'Connor, T.R.; Sikes, J.G.; Obradovic, Z.; Dunker, A.K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037–1049. [[CrossRef](#)]
112. Kim, J.H.; Lee, J.; Oh, B.; Kimm, K.; Koh, I. Prediction of phosphorylation sites using SVMs. *Bioinformatics* **2004**, *20*, 3179–3184. [[CrossRef](#)]
113. Koenig, M.; Grabe, N. Highly specific prediction of phosphorylation sites in proteins. *Bioinformatics* **2004**, *20*, 3620–3627. [[CrossRef](#)]
114. Zhou, F.-F.; Xue, Y.; Chen, G.L.; Yao, X. GPS: A novel group-based phosphorylation predicting and scoring method. *Biochem. Biophys. Res. Commun.* **2004**, *325*, 1443–1448. [[CrossRef](#)] [[PubMed](#)]
115. Huang, H.-D.; Lee, T.Y.; Tzeng, S.W.; Wu, L.C.; Horng, J.T.; Tsou, A.P.; Huang, K.T. Incorporating hidden Markov models for identifying protein kinase-specific phosphorylation sites. *J. Comput. Chem.* **2005**, *26*, 1032–1041. [[CrossRef](#)] [[PubMed](#)]
116. Xue, Y.; Li, A.; Wang, L.; Feng, H.; Yao, X. PPSP: Prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinform.* **2006**, *7*, 163.
117. Linding, R.; Jensen, L.J.; Ostheimer, G.J.; van Vugt, M.A.; Jørgensen, C.; Miron, I.M.; Diella, F.; Colwill, K.; Taylor, L.; Elder, K.; et al. Systematic discovery of in vivo phosphorylation networks. *Cell* **2007**, *129*, 1415–1426. [[CrossRef](#)]
118. Wong, Y.-H.; Lee, T.Y.; Liang, H.K.; Huang, C.M.; Wang, T.Y.; Yang, Y.H.; Chu, C.H.; Huang, H.D.; Ko, M.T.; Hwang, J.K. KinasePhos 2.0: A web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res.* **2007**, *35*, W588–W594. [[CrossRef](#)]
119. Plewczyński, D.; Tkacz, A.; Wyrwicz, L.; Rychlewski, L. AutoMotif Server for prediction of phosphorylation sites in proteins using support vector machine: 2007 update. *J. Mol. Model* **2008**, *14*, 69–76. [[CrossRef](#)]
120. Heazlewood, J.L.; Durek, P.; Hummel, J.; Selbig, J.; Weckwerth, W.; Walther, D.; Schulze, W.X. PhosPhAt: A database of phosphorylation sites in Arabidopsis thaliana and a plant-specific phosphorylation site predictor. *Nucleic Acids Res.* **2008**, *36*, D1015–D1021. [[CrossRef](#)]
121. Li, T.; Li, F.; Zhang, X. Prediction of kinase-specific phosphorylation sites with sequence features by a log-odds ratio approach. *Proteins* **2008**, *70*, 404–414. [[CrossRef](#)]
122. Wan, J.; Kang, S.; Tang, C.; Yan, J.; Ren, Y.; Liu, J.; Gao, X.; Banerjee, A.; Ellis, L.B.M.; Li, T. Meta-prediction of phosphorylation sites with weighted voting and restricted grid search parameter selection. *Nucleic Acids Res.* **2008**, *36*, e22. [[CrossRef](#)]
123. Yoo, P.D.; Ho, Y.S.; Zhou, B.B.; Zomaya, A.Y. SiteSeek: Post-translational modification analysis using adaptive locality-effective kernel methods and new profiles. *BMC Bioinform.* **2008**, *9*, 272. [[CrossRef](#)]
124. Saunders, N.F.W. Predikin and PredikinDB: A computational framework for the prediction of protein kinase peptide specificity and an associated database of phosphorylation sites. *BMC Bioinform.* **2008**, *9*, 245. [[CrossRef](#)] [[PubMed](#)]
125. Xue, Y. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol. Cell. Proteom.* **2008**, *7*, 1598–1608. [[CrossRef](#)] [[PubMed](#)]
126. Dang, T.H. Prediction of kinase-specific phosphorylation sites using conditional random fields. *Bioinformatics* **2008**, *24*, 2857–2864. [[CrossRef](#)] [[PubMed](#)]

127. Durek, P.; Schudoma, C.; Weckwerth, W.; Selbig, J.; Walther, D. Detection and characterization of 3D-signature phosphorylation site motifs and their contribution towards improved phosphorylation site prediction in proteins. *BMC Bioinform.* **2009**, *10*, 117. [[CrossRef](#)]
128. Biswas, A.K.; Noman, N.; Sikder, A.R. Machine learning approach to predict protein phosphorylation sites by incorporating evolutionary information. *BMC Bioinform.* **2010**, *11*, 273. [[CrossRef](#)]
129. Sobolev, B. Functional classification of proteins based on projection of amino acid sequences: Application for prediction of protein kinase substrates. *BMC Bioinform.* **2010**, *11*, 313. [[CrossRef](#)]
130. Jung, I. PostMod: Sequence based prediction of kinase-specific phosphorylation sites with indirect relationship. *BMC Bioinform.* **2010**, *11*, S10. [[CrossRef](#)]
131. Xue, Y. GPS 2.1: Enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection. *Protein Eng. Des. Sel.* **2011**, *24*, 255–260. [[CrossRef](#)]
132. Gao, J.; Xu, D. The Musite open-source framework for phosphorylation-site prediction. *BMC Bioinform.* **2010**, *11*, S9. [[CrossRef](#)]
133. Aravind, S.; Pablo, T.; Vamsi, K.M.; Sayan, M.; Benjamin, L.E.; Michael, A.G.; Amanda, P.; Scott, L.P.; Todd, R.G.; Eric, S.L.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).