DECIPHERING HUMAN CYTOPLASMIC PROTEIN TYROSINE KINASE PHOSPHORYLATION SPECIFICITY IN YEAST

THOMAS GEORG CORWIN

BERLIN 2015

DISSERTATION

ZUR ERLANGUNG DES GRADES EINES

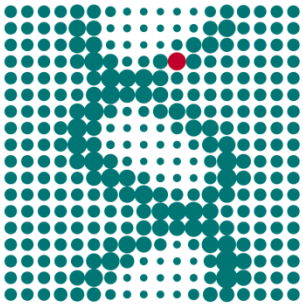
DOKTORS DER NATURWISSENSCHAFTEN

EINGEREICHT IM FACHBEREICH BIOLOGIE, CHEMIE, PHARMAZIE

DER FREIEN UNIVERSITÄT BERLIN

Max Planck Institut für molekulare Genetik

Otto Warburg Laboratory - Molecular Interaction Networks



MPIMG

MAX-PLANCK-GESELLSCHAFT

Freie Universität Berlin

Fachbereich Biologie, Chemie und Pharmazie

ERSTGUTACHTER: DR. ULRICH STELZL (MAX PLANCK INSTITUT FÜR MOLEKULARE GENETIK)
ZWEITGUTACHTER: PROF. DR. MARKUS WAHL (FREIE UNIVERSITÄT BERLIN)

TAG DER DISPUTATION: 05.11.2015

Zusammenfassung

Die Weitergabe von Signalen mittels Protein Tyrosin Kinasen (PTK) ist eine Eigenschaft von multizellulären Organismen und deren Deregulierung kann zur Entstehung komplexer Krankheiten wie Krebs im Menschen beitragen. Es gibt 90 humane PTKs unterteilt in 58 Rezeptor-PTKs und 32 nicht-Rezeptor PTKs (NRTKs). Es ist von essenzieller Bedeutung zu verstehen, wie jede einzelne NRTK ihre Substratproteine mit gegebener Spezifität erkennt und unterscheidet. Dies ist eine Herausforderung aufgrund überlappender Substratspezifität, enzymatischer Aktivität unterschiedlicher Größenordnungen und zelltypabhängiger Expressionsstärke zwischen NRTKs. *Saccharomyces cerevisiae* (Bäckerhefe) hat keine PTKs und es wurde gezeigt, dass aktive, menschliche NRTKs einzeln und schwach in Hefe exprimiert werden können ohne einen Hefe-Phänotyp zu erzeugen. Dabei werden Hefeproteine phosphoryliert, was mittels moderner massenspektrometrischer Methoden hintergrundfrei gemessen werden kann. Die Hälfte aller 32 NRTKs, mit Repräsentanten aus allen NRTK Familien, zeigten Aktivität in Hefe. Durch 60 massenspektrometrische Messungen in einem einzelnen Versuchsaufbau, wurden 1433 Phosphorylierungsstellen auf 900 Hefeproteinen gemessen. Die Messungen ermöglichten die Analyse der überlappenden Spezifität zwischen NRTKs und die Definierung linearer Aminosäuresequenz-Motive für jede einzelne NRTK, welche die Zuordnung von NRTKs zu 1388 aus 13240 bekannten menschlichen Phosphorylierungsstellen ermöglichte. Die Konservierung von Modifizierungsstellen zwischen Mensch und Hefe ermöglichte die Zuordnung von NRTKs zu 63 orthologen Modifizierungsstellen und führte zur Entdeckung, dass (aerobe) Glykolyse und Signalwege onkogener NRTKs enger verknüpft sind als bisher angenommen. Für NRTKs wurde individuelles und gemeinsames Modifizieren von einzelnen Substratproteinen und Substratfamilien vorhergesagt, was exemplarisch mittels eines *in vitro* Kinasen-Testverfahren experimentell überprüft wurde. Die Dechiffrierung menschlicher NRTK-Spezifität in Hefe resultierte in einem neuen Datensatz, welcher die Basis für weitere experimentelle Analysen sein könnte, um die systemweite Signalweitergabe von NRTKs in normalwachsenden und wuchernden Zellen zu entschlüsseln.

Schlagwörter:

Protein Tyrosine Kinase, PTK, zytoplasmisch, NRTK, homo sapiens, Mensch, Saccharomyces cerevisiae, Hefe, Spezifität, lineares Sequenzmotiv, Motiv Scoring, Protein-Protein Interaktions Netzwerk, PPI Netzwerk

Abstract

Protein tyrosine kinase (PTK) signaling can be regarded as a hallmark of multi-cellular organisms and its deregulation causes a variety of complex diseases in human including cancer. Therefore, it is essential to understand how each individual tyrosine phosphorylating enzyme is able to recognize and distinguish its target proteins. This task still remains challenging due to overlapping substrate specificity, magnitude differences in enzymatic activity, and cell type dependent expression levels of PTKs. There are 90 human PTKs with 58 receptor-type PTKs and 32 non-receptor type PTKs (NRTKs) thereof. *Saccharomyces cerevisiae* is lacking PTKs and it was shown that, when individually expressed at low levels, active human NRTKs phosphorylate yeast proteins - in a background-free environment, in intact cells with high specificity and without a growth phenotype. Half of all 32 NRTKs representing all NRTK families showed activity in yeast. Hence, individual NRTK specificity on both linear amino acid motif- and structural level was determined by 60 measurements in a single experimental set-up recording 1433 tyrosine phosphorylation sites on 900 proteins in yeast using state-of-the-art phospho-proteomics. The mass-spectrometric measurements enabled analysis of NRTK specificity overlap and the determination of linear amino acid sequence motifs for individual NRTKs. Motif-based scoring of 13240 reported human pY-sites enabled kinase inference for 1388 pY-sites. Modification site conservation between yeast and human enabled kinase inference for 63 orthologous pY-sites and led to the discovery that aerobic glycolysis and oncogenic NRTK signaling may be more highly inter-linked than previously assumed. Individual and common NRTK targeting of single and families of substrate proteins was predicted and exemplary experimentally verified by an *in vitro* kinase assay. Deciphering human NRTK specificity in yeast provided a novel data set which may aid further experimental analysis to unravel system-wide NRTK signaling in normal and proliferating cells.

Keywords:

Protein Tyrosine Kinase, PTK, NRTK, cytoplasmic, homo sapiens, human, Saccharomyces cerevisiae, yeast, specificity, linear sequence motif, motif scoring, Protein-protein interaction network, PPI network

Contents

Abbreviations

ATP - Adenosine triphosphate

CC - Clustering coefficient

C-terminal - carboxyl-terminal

FDR - False Discovery Rate

FERM - band 4.1, ezrin, radixin, moesin homology

FP - False Positive

FRET - Förster resonance energy transfer

GEC - Glycolytic enzyme complex

GEF - Guanine nucleotide exchange factors

GO - Gene ontology

HPLC - high performance liquid chromatography

HPR - Horse Radish Peroxidase

Hs - *Homo sapiens*

IBC - In-betweeness centrality

IMAC - Immobilized metal-affinity chromatography

KSRs - Kinase-substrate relationships

LC - Liquid chromatography

MHR - MIG6 homology region

MOAC - Metal oxide affinity chromatography

MS - Mass spectrometry

NLS - Nuclear localization sequence

NRTK - non-receptor type (cytoplasmic) tyrosine kinase

N-terminal - amino-terminal

ORF - Open reading frame

PAGE - Polyacrylamide gel electrophoresis

PCR - Polymerase chain reaction

PH domain	-	pleckstrin homology domain
PK	-	Protein kinase
PPI	-	Protein-protein interaction
PPM	-	metallo-dependent protein phosphatases
PPP	-	phosphoprotein phosphatase
PSSM	-	Position-specific scoring matrix
PTB	-	Phosphotyrosine-binding
PTK	-	Protein tyrosine kinase
PTM	-	Post-translational modification
PTP	-	phosphotyrosine phosphatases
pY	-	Phospho-tyrosine
pY-site	-	tyrosine phosphorylation site
RPTK	-	Receptor protein tyrosine kinases
RT	-	room-temperature
Sc	-	*Saccharomyces cerevisiae*
SFK	-	SRC-family kinases
SH2	-	Src-homology 2 (domain)
SH3	-	Src-homology 3 (domain)
SILAC	-	Stable isotope labeling of amino acids in cell culture
TCA	-	tricarboxylic acid
TOF	-	Time-of-flight
TP	-	True positive
Y-site	-	(15mer) 7 residues flanking a tyrosine residue

1. Introduction

1.1. Protein phosphorylation

Cells of all organisms store and transmit information via post-translational modification (PTM) of proteins. One of the most prominent PTMs is reversible phosphorylation of proteins on the hydroxl-group of three hydroxyamino acids: serine, threonine, and tyrosine. Phosphate has ideal properties for the generation of biological molecules: it is highly abundant in nature and it has chemical advantages such as high water solubility and the ability to form chemical linkages that are stable in water at moderate temperatures (Westheimer, 1987). Thus, phosphate ester linkages were employed early in evolution for the formation of the nucleic acids and later to reversibly modify the otherwise genetically fixed 20 natural amino acids (Westheimer, 1987). As such, new chemical entities are created that diversify biochemical properties of proteins by addition of negative charge to the protein surface. The universal phosphate donor and enzyme catalyst in biological systems is the highly abundant adenosine triphosphate (ATP) which is converted into adenosine diphosphate (ADP) by donating an orthophosphate molecule for a phosphorylation reaction, which in turn results in an overall release of energy to the system (Hunter, 2012). Serine, threonine and tyrosine residues are the amino acids predominantly phosphorylated in higher eukaryotes with an estimated phosphorylation distribution in humans of approximately 88, 11, and 1 percent, respectively (Olsen et al., 2006, Junger and Aebersold, 2014). Six other amino acids, arginine, glutamine, aspartate, cysteine, histidine, and lysine, are also known to be phosphorylated (Hunter, 2012). Phosphorylation of the latter amino acids is less stable and hence can be easily missed in proteomic approaches employing unfavorable conditions such as the use of acids. Histidine phosphorylation, for example, is by far less studied than tyrosine phosphorylation even though histidine modification was estimated to be more frequent than modification of tyrosine (Ciesla et al., 2011).

1.2. Regulation of protein phosphorylation

Cellular signaling via phosphorylation is regulated by a three-tier system employing three classes of proteins with modular domain architecture that can either add (“writers”) or recognize (“readers”) or remove (“erasers”) phosphate from substrate proteins as shown in Figure 1. There are enzymes with specificity for serine and threonine residues (serine/threonine (S/T) kinases, -binding domains, and -phosphatases) and others specific for tyrosine residues (tyrosine kinases, -binding domains, and -phosphatases). Furthermore, dual-specific enzymes which can phosphorylate, recognize, or de-phosphorylate both have been described (Dhanasekaran and Premkumar Reddy, 1998, Becker and Joost, 1999, Camps et al., 2000).

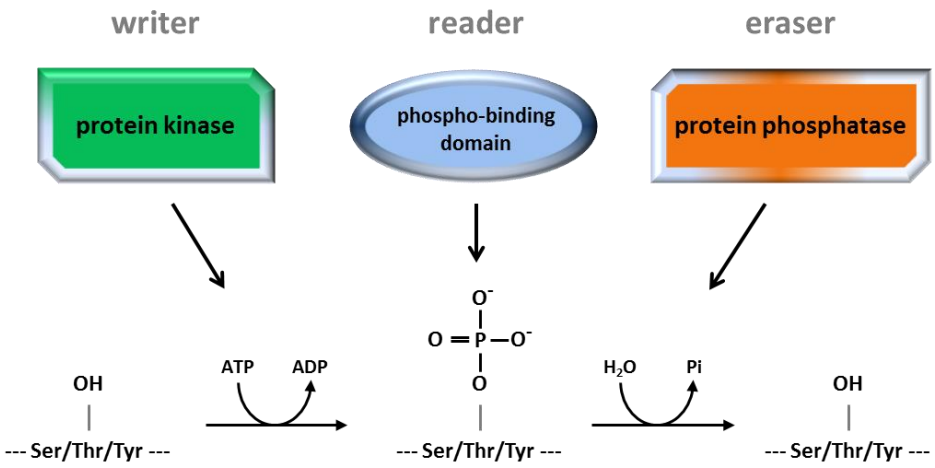


Figure 1: Schematic view of the three-part system of phosphorylation based cellular signaling. Protein kinases ("writers") catalyze the phosphorylation mainly of serine, threonine, and tyrosine residues under ATP consumption. Proteins containing phosphorylation binding domains ("reader") bind to the modified residues. Protein phosphatases ("erasers") remove phosphate by hydrolysis and released pyrophosphate is recycled by the cell. Modified from Jin and Pawson (2012).

Human genome sequencing enabled the identification of 518 protein kinases (PK) which constitute the class of "writers" (Manning et al., 2002b). These enzymes catalyze the phosphorylation reaction and hence regulate most cellular processes including metabolism, transcription, cell cycle progression, apoptosis, differentiation, cytoskeletal rearrangement, and cell mobility (Manning et al., 2002b). In addition to PKs, approximately 150 human protein phosphatase catalytic subunits were identified (Alonso et al., 2004) which can be divided into three main families of "erasers": the phospho-protein phosphatase (PPP) family, the metallo-dependent protein phosphatase (PPM) family, and the phospho-tyrosine phosphatase (PTP) family (Moorhead et al., 2009). Along with approximately 90 protein tyrosine kinases (PTK), 107 genes were identified in the human genome encoding four families of PTPs (Alonso et al., 2004). While phosphatases are not as highly investigated as PKs, it is increasingly appreciated that they play an equally important role in steering cellular processes enabling regulated reversibility of the PTM. Indeed, the phosphorylation status of an amino acid is dictated by an equilibrium of attachment and removal of phosphate by the appropriate enzymes.

Signaling molecules are constructed from domains in a modular fashion which determine molecular interactions with nucleic acids or other proteins, or have enzymatic activity (Jin and Pawson, 2012). Interaction surfaces on proteins can be recognized by interaction domains which bind, for example, to specific peptide stretches defined by amino acid arrangement or modification, or higher-order structure. By recognition of PTMs or chemical second messengers, interaction domains can target a protein to a specific subcellular localization, initiate protein signaling complex formation, and control the conformation, activity, and substrate specificity of enzymes (Pawson and Nash, 2003). Different

interaction domains have evolved to bind to specific proteins, phospho-lipids, or nucleic acids. Proteins are targeted by domain recognition of one of more properties such as PTMs and specific peptide sequences, and via homo- or heterotypic domain-domain interactions (Pawson and Nash, 2003). Important representatives of binding domains are the SRC-homology 2 (SH2) and phospho-tyrosine binding (PTB) domains interacting with phosphorylated tyrosine (Jin and Pawson, 2012, Grossmann et al., 2015). Furthermore, 14-3-3 domains and WW domains bind to phosphorylated serine, threonine, and poly-proline peptide recognizing SRC-homology 3 (SH3) domains while PDZ domains bind both to specific peptide sequences on protein carboxyl-terminals and to other PDZ domains, for instance (Pawson and Nash, 2003). Regulation of cellular signaling by phosphatases and modular binding domains are research fields in its own right whereas this thesis focuses on the determination of substrate specificity of PTKs belonging to the group of "writers".

1.3. Protein tyrosine kinases

Serine and threonine phosphorylation can be found in all eukaryotes whereas tyrosine phosphorylation evolved with the emergence of multi-cellularity (Suga et al., 2012). As mentioned above, around 90 genes were identified encoding PTKs in the human genome by mapping tyrosine kinase domain sequence conservation (Manning et al., 2002b). 58 cell membrane-spanning receptor protein tyrosine kinases (RPTK) were classified into 20 subfamilies and 32 non-receptor tyrosine kinases (NRTKs) were distributed to 10 subfamilies by phylogenetic analysis of amino acid sequences and domain architecture (Robinson et al., 2000). PKs, and in particular PTKs, constitute a large fraction of all known tumor-suppressor and dominant oncogenes whereat PTKs represent the largest group of dominant oncogenes with structural homology (Blume-Jensen and Hunter, 2001). Even though PTKs comprise only 0.3 percent of all metazoan genes, somatic mutations in these genes cause a significant fraction of human cancers (Blume-Jensen and Hunter, 2001). In general, PTK signaling needs to be precisely coordinated and integrated at all times during embryonic development and adult life by allostery and protein-protein interactions to ensure proper protein function and prevent oncogenesis (Blume-Jensen and Hunter, 2001, Jin and Pawson, 2012). Even though the overall fold of PTKs is conserved, fine-tuned differences in the core sequence and flanking regions of kinases allow each kinase to respond to a unique set of signals resulting in their activation and inhibition (Nolen et al., 2004).

The majority of 58 RPTKs are activated via dimerization at the cell membrane upon extra-cellular stimulation and subsequent phosphorylation of their cytoplasmic tails in *trans* by the dimerization partner (Schlessinger, 2000). Phosphorylation of receptor tails enables interaction with binding domains of downstream signal transducers including other kinases such as cytoplasmic PTKs as

reviewed by Joseph Schlessinger (2014) - a pioneer in the field of RPTKs. As RPTKs are not in the focus of this thesis only NRTKs and their regulation are described in more detail below.

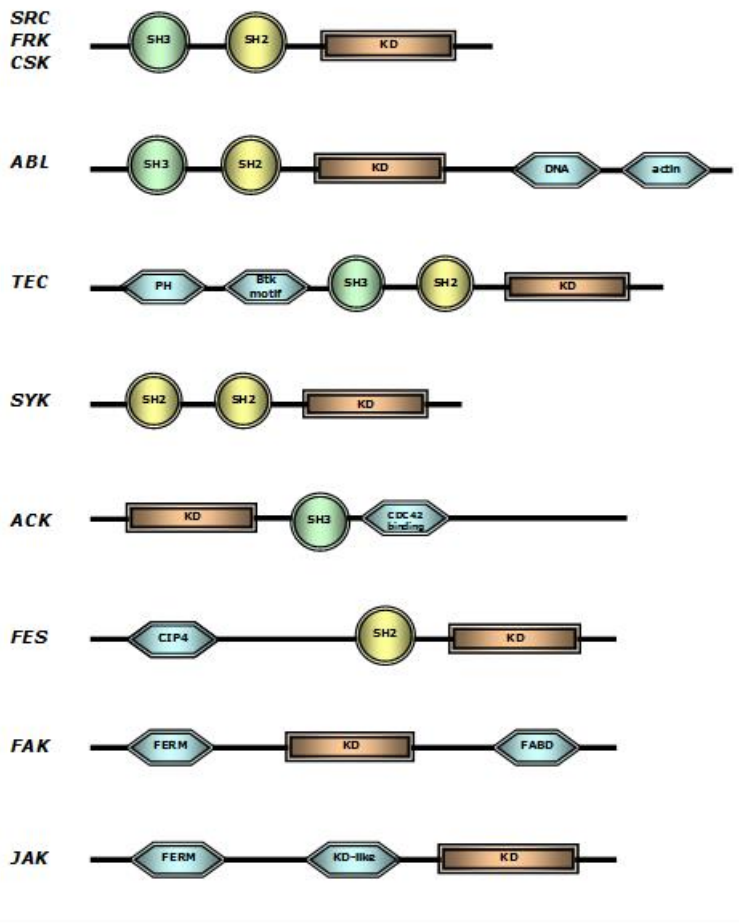


Figure 2: Human cytoplasmic protein tyrosine kinase (NRTK) families assigned according to their domain architecture (modified from Blume-Jensen and Hunter (2001)). KD = kinase domain.

Cytoplasmic or non-receptor type PTKs are constructed of modular domains which are regulating substrate recruitment and the enzymes' catalytic activity by conformational switches upon post-translational modification. These phosphorylating enzymes can be grouped into 10 families according to their structural similarity (Figure 2). A viral version of cellular SRC, therein called v-SRC, was identified as a gene of Rous sarcoma virus responsible for viral cell-transforming ability and was the first identified tyrosine kinase (Hunter and Sefton, 1980). There are eight SRC-family kinases (SFKs) which share a common domain architecture comprising a conserved carboxyl-terminal (C-terminal) kinase domain preceded by one SH2 domain and one SH3 domain (Figure 2). SFKs play a key role in coupling receptor signals with the cytoplasmic signaling machinery by many molecular strategies (Parsons and Parsons, 2004). Fundamental cellular processes influenced by SFK signaling include cell growth, cell differentiation, cell shape, migration, and survival (Parsons and Parsons, 2004). A C-terminal tyrosine (Y527) is conserved among SFKs and is known to be phosphorylated by c-SRC tyrosine kinase (CSK) and other NRTKs (Roskoski, 2004, Boggon and Eck, 2004). Y527 phosphorylation

keeps SFKs in an inactive state by intra-molecular binding of their SH2 domain to the phosphorylated tail as confirmed by the first reported SFK crystal structures (Sicheri and Kuriyan, 1997). As an additional intra-molecular inhibitory mechanism, the SH3 domain of SFKs interacts with the linker region between the SH2 domain and the amino-terminal (N-terminal) kinase lobe and thus is packed against the back of the kinase domain (Brown and Cooper, 1996, Bradshaw, 2010). Most, but not all, protein kinases are activated by phosphorylation in an activation segment causing conformational change whereas a single tyrosine phosphorylation (pY416 in chicken SFKs) activates SFKs (Johnson et al., 1996, Nolen et al., 2004). Abelson kinases, ABL1 and ABL2 (ARG), are closely related to SFKs yet lack the C-terminal inhibitory tyrosine 527 (Nagar et al., 2003). The N-terminal half of ABL kinases resemble closely SFKs with the exception of a N-terminal myristoylated "cap" unique for ABL kinases which can bind the kinase domain (Nagar et al., 2003). This conformational change enables SH2 and SH3 domain binding to the ABL kinase domain and linker region similar to SFKs however, in a distinct manner. Intra-molecular domain binding inactivates ABL kinase whereas the ABL kinase domain is kept in an open conformation as observed for active SFKs (Nagar et al., 2003). The FRK (FYN-related kinase or Nuclear tyrosine protein kinase RAK) family of NRTKs is also related to SFKs and consists of three members: FRK, Tyrosine-protein kinase 6 (PTK6; also known as Breast Tumor Kinase, BRK), and SRMS (SRC-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristoylation sites). Mutational and mass spectrometry analysis on PTK6 suggested that Y341 is an activating auto-phosphorylation site and that C-terminal Y447 has an auto-inhibitory function analogous to the corresponding regulatory sites Y416 and Y527 in SFKs (Qiu and Miller, 2002). Structural analysis of the SH2 domain of PTK6 bound to a peptide derived from the C-terminus of PTK6 harboring Y447 supported the suggested auto-inhibition involving Y447 and revealed a PTK6-specific SH2 domain fold (Hong et al., 2004). Moreover, it was shown that the SH3 domain of PTK6 interacts with a linker region between SH2 and kinase domains via hydrophobic residues comprising a PTK6-specific auto-inhibitory mechanism (Ko et al., 2009). TEC (tyrosine kinase expressed in hepatocellular carcinoma) family NRTKs are primarily expressed in hematopoietic cells where they function downstream of several immune cell receptors (Joseph and Andreotti, 2009, Schwartzberg et al., 2005). There are five TEC family members including BMX (bone-marrow tyrosine kinase gene on chromosome X; also known as ETK) with a domain architecture similar to SRC and FRK family kinases with a single N-terminal SH3 domain followed by a single SH2 domain and kinase domain (Blume-Jensen and Hunter, 2001). Unlike SFKs, TEC kinases (with the exception of BMX) also contain a pleckstrin homology (PH) domain followed by a TEC homology (TH) domain harboring one or two proline-rich regions at the N-terminus (Schwartzberg et al., 2005). Furthermore, TEC kinases are lacking the inhibitory C-terminal tyrosine residue (Schwartzberg et al., 2005). Activation of TEC family tyrosine kinases occurs downstream of both SRC and SYK family kinases in corresponding signaling pathways (Bradshaw,

2010). Activation of TEC family kinases requires (A) recruitment to the plasma membrane via their PH domains interacting with phosphatidylinositol (3,4,5)-trisphosphate (PIP3) created by the activation of PI3'-kinase and/or other proteins, (B) phosphorylation by SFKs or auto-phosphorylation, and (C) interactions with other proteins that bring the TEC-family kinases into antigen-receptor signaling complexes (Schwartzberg et al., 2005, Bradshaw, 2010). Furthermore, TEC family kinases are regulated by intra- and inter-molecular interactions between SH3 domain and proline-rich regions and between the SH2 domain and another molecule's SH3 domain (Schwartzberg et al., 2005, Bradshaw, 2010). Isomerization of a proline residue in the SH2 domain by Peptidylprolyl isomerase A changes its specificity from TEC SH3 domain recognition towards interactions with other proteins (Schwartzberg et al., 2005, Bradshaw, 2010). SYK (spleen tyrosine kinase) and closely related ZAP-70 (zeta-chain (TCR) associated protein kinase 70kDa) constitute another type of NRTK family. They function downstream of antigen receptors and SFK signaling in immune cells and they are composed of two N-terminal SH2 domains followed by an inter-domain linker and a C-terminal kinase domain (Gradler et al., 2013). A linker between the SH2 domains serves as docking surface for immune receptor tyrosine-based activating motifs (ITAMs) that are displayed at the cytosolic side of the plasma membrane (Siraganian et al., 2002). The ITAM motif is phosphorylated during immune receptor signaling and can consequently bind and activate SYK whereas ZAP-70 requires additional phosphorylation by the SFK LCK (Siraganian et al., 2002, Gradler et al., 2013). In addition, phosphorylation of several tyrosine residues in the activation loop and linker regions activates SYK by causing conformational changes in the kinase domain (Gradler et al., 2013). The ACK family of NRTKs has also two members, ACK1 (TNK2) and TNK1, which are unique among NRTKs having one N-terminal SAM (Sterile Alpha Motif) domain and one SH3 domain C-terminal to the kinase domain followed by a MIG6 homology region (MHR). Tyrosine 284 in the ACK1 activation loop was shown to be targeted by SFKs whereas auto-phosphorylation increases the, compared to other NRTKs, low activity only marginally (Gajiwala et al., 2013). Structural analysis suggested that ACK1 is auto-inhibited in monomeric form and activated by symmetric dimerization mediated by the SAM domain (Gajiwala et al., 2013). The regulatory function of the SH3 domain remains elusive whereas an influence on ACK1 dimerization and SH3 domain interaction with the MHR orienting the MHR for inhibitory interactions with the kinase domain was proposed (Gajiwala et al., 2013). The FES family of NRTKs also consists of two members, FES (Feline sarcoma/Fujinami avian sarcoma oncogene homolog) and FER (Fujinami poultry sarcoma/Feline sarcoma-related protein Fer). FES has been implicated in cytokine signal transduction, hematopoiesis, and embryonic development (Cheng et al., 1999). As other NRTKs FES and FER harbor an auto-phosphorylation site (Y713) to which the only SH2 domain was shown to bind and may influence their kinase activity (Hjermstad et al., 1993). The unique N-terminal sequence of FES contains two regions with strong homology to coiled-coil forming

domains and it was suggested that FES activity is largely dependent on oligomerization of the kinase regulated by coiled-coil interactions (Cheng et al., 1999). Mutations in the first coiled-coil region in FES resulted in enhanced kinase activity. However, mutations in either the first or second coiled-coil motif of FER abolished oligomerization however, had no effect on auto-phosphorylation activity (Cheng et al., 1999, Greer, 2002). Moreover, a version of FER lacking the coiled-coil region exists as a fully active monomer (Greer, 2002). Thus, there are substantial differences in the regulation of FER and FES whereas FES may be under tighter control requiring oligomerization via coiled-coil interactions (Greer, 2002). Having a N-terminal FERM (band 4.1, ezrin, radixin, moesin homology) domain followed by a linker region, a kinase domain, a proline-rich low-complexity region, and C-terminal focal adhesion targeting (FAT) domain, the two NRTKs focal adhesion kinase (FAK or PTK2) and proline-rich tyrosine kinase 2 (PYK2) constitute another NRTK family namely FAK. FAK is auto-phosphorylated at tyrosine 397 in the linker region between the FERM and kinase domains upon integrin and growth factor activation (Schaller et al., 1994). Phosphorylated Y397 and poly-proline motif (PxxP), which are in close proximity, recruit and activate SRC which binds via its SH2 and SH3 domains (Lietha et al., 2007). SRC in turn phosphorylates Y576 and Y577 in the activation loop of FAK resulting in its full activation (Calalb et al., 1995). By solving a structure of a large fragment of FAK including the FERM and kinase domain in inactive form, Lietha et al. (2007) could show that the auto-inhibited assembly is stabilized through an interaction between the FERM domain and the kinase C-lobe which blocks the kinase domain directly and does not displace the C-helix from the active site as observed for SFKs. The authors could further show that the FERM domain protects FAK from Y397 phosphorylation and from activation segment targeting by SRC. In addition to kinase activity, SH2, SH3, and other domains can also regulate kinase specificity by recruitment of selected substrates to the catalytic site (Pawson and Nash, 2003).

1.4. Determination of NRTK substrates and specificity

In order to understand PTK signaling it is necessary to define which proteins are phosphorylated by PTKs and how these substrates are recognized. Phosphorylation is the most highly studied PTM and emerging high-throughput techniques enabled identification of 203997 human phosphorylation events (phosphositeplus.org dataset 06/04/2015). Indeed, it was estimated that the percentage of phospho-proteins in eukaryotic proteomes is in the range of 40 to 45 percent or even higher (Junger and Aebersold, 2014). However, only approximately six percent of all reported human substrates and modification sites (approximately 4 percent of all reported tyrosine phosphorylation sites) are linked to responsible protein kinases (phosphositeplus.org dataset 06/04/2015). A major challenge is that at a given point in time inside each cell, PTKs are differentially expressed dependent on cell type or cell cycle phase, and show partly overlapping specificity and orders of magnitude differences in

enzymatic activity (Bewarder et al., 1996, Huttlin et al., 2010). Thus, attempts to measure the complete phosphorylation status of a single cell or cells without knowing the responsible enzymes represent only one side of the coin in constructing and understanding system-wide phosphorylation networks and depict only a snap-shot of the dynamics in cellular signaling (Newman et al., 2014). In this regard, established databases cataloging phosphorylation sites, kinase-substrate relationships (KSRs), and kinase specificity determinants are useful however, far from being complete. The following part of this introduction focuses on methodologies to determine that a protein *is,* or *can be,* phosphorylated by a kinase. In addition to defining kinase substrates, understanding phosphorylation-based signaling dynamics also concerns information about the *when* and *where*. Methodologies to gain insight into the when and where such as Förster resonance energy transfer (FRET)-based systems have been developed which were recently reviewed by Newman et al. (2014).

A prerequisite of enzymes to act on phosphorylation sites is the co-expression of both enzyme and substrate in the same subcellular localization and that they are physically interacting. The presence and abundance of proteins in a cell is regulated on a transcriptional and on a post-transcriptional level which can be monitored by complementary assays. For instance, gene expression and mRNA transcript levels can be determined by quantitative real-time polymerase chain reaction (qPCR), expression microarrays, or RNA sequencing. Transcriptional activation of a promoter can be measured using reporter gene assays and changes in protein abundance can be monitored via Western blotting, fluorescent imaging techniques, and quantitative mass-spectrometry based approaches (Newman et al., 2014, Junger and Aebersold, 2014). Testing when and where protein-protein interactions (PPI) occur was traditionally conducted using biochemical assays employing subcellular fractionation techniques or co-immuno-precipitation by epitope-tags, interaction domains, or antibodies followed by Western blotting. Using a tandem affinity tag (TAP) genetically fused to all proteins or proteins-of-interest enabled immuno-precipitation of protein complexes in yeast on genome-wide scale (Gavin et al., 2006, Krogan et al., 2006). This strategy was also applied to sets of human phosphatases and kinases (Wepf et al., 2009, Varjosalo et al., 2013). However, the methodology is limited in capturing transient PPIs which are lost upon cell lysis and/or in the absence of cross-linking agents. Furthermore, the data may not provide direct evidence for PPIs to occur as they may take place in the context of multimeric protein complexes (Newman et al., 2014). In order to overcome some of these limitation yeast-two-hybrid screening has been performed to define candidate substrates having the advantage to also capture transient PPIs such as kinase-substrate interactions in a high-throughput manner (Stelzl et al., 2005, Parrish et al., 2006). In yeast-two-hybrid screening the kinase can be used as a “bait” or being expressed in addition to “bait” and “prey” in order to test for phosphorylation dependent interactions (Yang et al., 1992, Worseck et al., 2012, Grossmann et al., 2015). Moreover, FRET-based assays and protein complementation assays (PCA)

(or a combination thereof) which rely on the interaction-dependent reassembly of a split reporter protein were used to study protein-protein interactions, both qualitatively and quantitatively (Shekhawat and Ghosh, 2011). More recently, fluorescence imaging techniques such as immuno-fluorescence and live cell imaging have been developed with continuously increasing resolution to define the subcellular location and interactions between enzymes and substrates in their native environment (Dedecker et al., 2012).

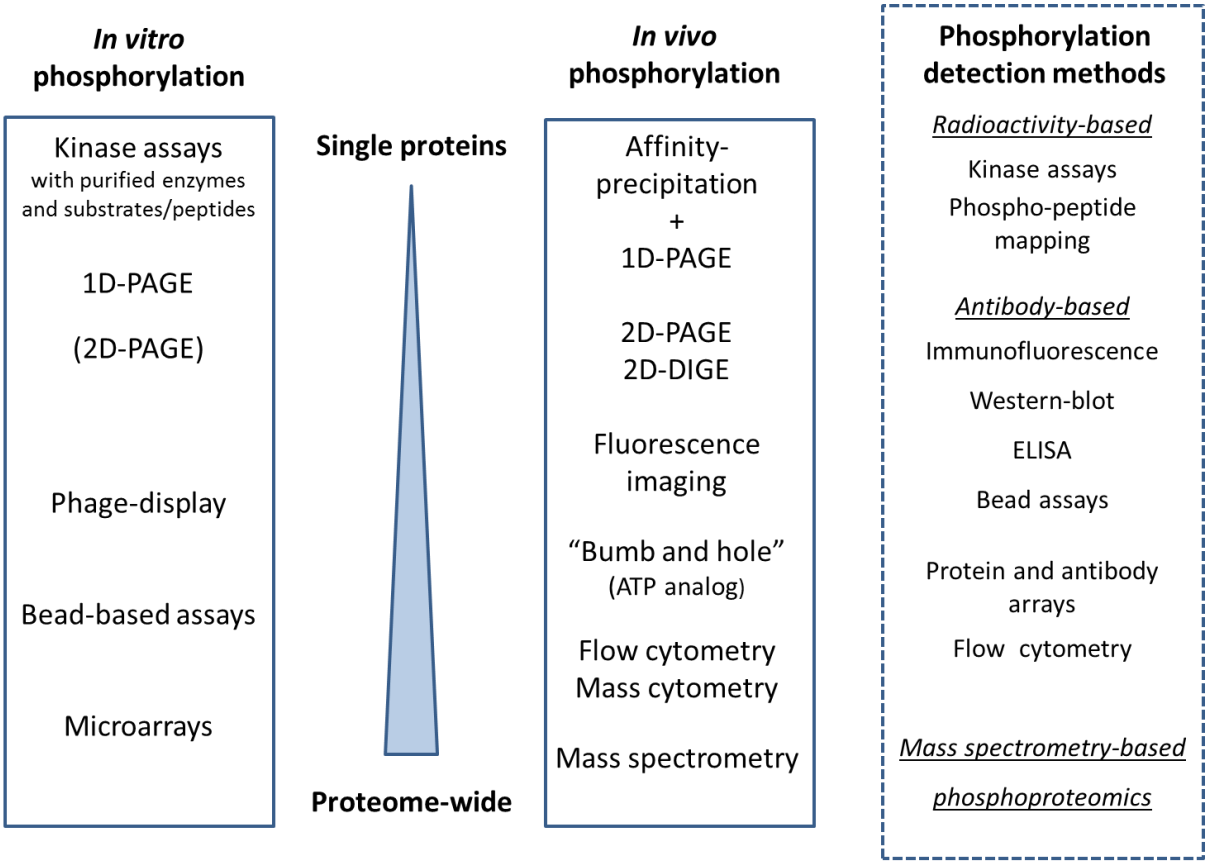


Figure 3: Overview methods for the determination and detection of *in vitro* and *in vivo* phosphorylation sites. Methods are ordered vertically from low- to high-throughput and the box on the left side indicate strategies for the detection of protein phosphorylation as summarized by Junger and Aebersold (2014). Typically, a combination of these methods is employed for the characterization of kinase-substrate relationships (KSRs).

Methodologies developed to detect phosphorylation events and to define kinase-substrate relationships (KSRs) linking substrates to responsible kinases, can be split into approaches which begin at the protein level and work their way to the cellular systems level and into approaches working in the inverse direction as indicated in Figure 3. In regard of methods analyzing the biochemical properties of candidate substrate on a protein level many kinase-substrate interactions require scaffolding proteins, second messengers, and other co-factors to occur which may not be provided in an *in vitro* situation. Hence, all *in vitro* methodologies presented here require subsequent validation assays in an appropriate cellular system to ensure that the kinase reaction can actually

occur under physiological conditions. Different types of methodologies for kinase substrate identification and databases for indexing the resulting information are presented.

1.4.1. In-vitro kinase assays: from single to thousand proteins

In order to test whether a kinase can phosphorylate a candidate substrate most commonly an *in vitro* kinase assay is conducted. Traditionally, a single purified kinase was incubated with a single purified candidate substrate in the presence of ATP and the expected phosphorylation (or expected abolished phosphorylation by site-directed mutagenesis) was subsequently detected via autoradiography or scintillation counting using radiolabeled (gammaP32) ATP. Due to safety and cost issues this modification read-out was replaced by non-radioactive assays employing enzymatic reactions to measure ATP depletion or ADP production, or by Western blotting with epitope-of-interest specific antibodies subsequent to polyacrylamide gel electrophoresis (PAGE) (Newman et al., 2014). This method commonly termed kinase assay has been a valuable tool to confirm and functionally characterize candidate substrates however, it is tedious and time-consuming. Thus, this methodology was scaled up by the development of high-throughput techniques such as phage display or protein microarrays. For example, expression of a cDNA library in bacteriophage led to the discovery of seven SRC substrates by incubating the phage display with extracts containing baculovirally expressed human SRC kinase and ATP followed by substrate identification using a tyrosine phosphorylation-specific antibody (Lock et al., 1998). Later, functional protein microarrays were generated by printing thousands of recombinantly expressed and purified proteins, or sets of purified candidate substrates, on a microarray and floating the array with purified kinases together with ATP (Zhu et al., 2001, Ptacek et al., 2005, Newman et al., 2013). After some washing steps, phosphorylation events can be determined by the position on the array typically by employing radiolabeled ATP or a fluorescent detector linked to an epitope- or modification-specific antibody. Newman et al. (2013) incubated individual protein microarrays containing 4191 unique, full-length human proteins with 289 active, full-length human kinases together with radiolabeled ATP which identified 24046 *in vitro* phosphorylation events on 1967 substrates. This dataset was refined by applying a Bayesian statistics model where the likelihood of KSRs was determined in terms of physical interactions, co-localization, and co-expression of kinase and putative substrates resulting in 3656 refined KSRs. By integration of reported MS determined *in vivo* phosphorylation site sequences, amino acid sequence motifs were generated. This refined, probabilistic KSR dataset was used together with the obtained sequence motifs and reported *in vivo* modification sites to construct a high-resolution map of human phosphorylation networks linking 4417 phospho-sites on substrates to their cognate kinase (Newman et al., 2013). In this way, almost the entire proteome can be queried in a single experiment *in vitro*. However, due to high background signals many transient KSRs may be missed as their signal may be

below a necessary high signal-to-noise cut-off. Other confounding factors are the substantial variation of protein levels on the chip or misfolding and truncation of substrates during purification as well as contamination with undefined kinase activities acquired in target kinase purification (Newman et al., 2014). In order to better mimic to the cellular environment, cell lysates were floated over the array together with the kinase. In this way, co-factors are available for *in vitro* kinase reactions however, several other kinases are simultaneously present and active in cell extracts at given point in time. Thus, there is strong background phosphorylation when trying to identify responsible kinases for proteome-wide phosphorylation changes. Nevertheless, this method allowed identification of changes in the phospho-proteome by using differentially conditioned cell lysates and it is applicable for biomarker identification (Woodard et al., 2013). Another variation of a functional protein microarray is the reverse phase protein microarray where cell lysate is spotted on a glass surface, for example, and an antibody specific for a tag or PTM is applied to the array. This enables comparative quantification of proteins-of-interest which is useful to economically analyze diverse clinical samples, for instance (Ummanni et al., 2014). Nevertheless, kinase substrate relationships remain elusive. Strategies to capture *in vivo* phosphorylation events are presented in the following section which can also define the phosphorylation state of a cell. However, almost all methods are limited in defining kinase-substrate pairs due to strong cellular kinase activity as mentioned above.

1.4.2. System-wide characterization of *in vivo* phosphorylation events

Initially, PTK substrates were identified by screening for phosphorylated proteins from radio-labeled cells over-expressing PTKs or upon growth factor stimulation of RPTKs by employing two-dimensional PAGE (2D-PAGE) (Hunter and Sefton, 1980). Radioactive phosphate is incorporated into proteins by the PTK and cell lysate is applied to a polyacrylamide gel. Then, proteins are separated by both iso-electric focusing and electrophoresis on a single gel in two dimensions. Subsequently, phosphorylated proteins are detected and identified by applying a photographic film to the gel which darkens at the positions of modified proteins. Alternatively, 2D-PAGE separated proteins and protein complexes were stained using a Coomassie dye or silver-stain, subsequently excised from the gel and analyzed using mass spectrometry (Camacho-Carvajal et al., 2004). An extension of 2D-PAGE is the use of spectrally resolvable, size and charge-matched fluorescent dyes for labeling multiple cell extracts which hence can be simultaneously applied and quantitatively compared on a single gel (Marouga et al., 2005). This method was termed two-dimensional differential gel electrophoresis (2D-DIGE). Even though 2D-PAGE based methods such as 2D-DIGE allowed identification of hundreds of modified proteins in parallel it can capture only a fraction of thousands of modifications on thousands of proteins in a cell at a given time-point.

In approaches employing one-dimensional PAGE target proteins are typically analyzed by Western blotting or by migration-pattern on the gel. The quality of Western blotting read-out strongly depends on the specificity of the antibody used. For more quantitative analysis an enzyme-linked immuno-sorbent assay (ELISA) set-up may be employed (Junger and Aebersold, 2014). If a phospho-specific antibody is not available there are general phospho-detection reagents including Pro-Q Diamond stain (LifeTechnologies, Inc.) or the Phos-Tag phospho-chelator that binds specifically phosphorylated ions (Manac, Inc.) (Newman et al., 2014). In the case of global phospho-tyrosine recognition very sensitive antibodies have been developed into major research tools - in particular the pan pY-antibody 4G10 (Kanakura et al., 1990), the pan pY-antibody P-Tyr-100 (Ross et al., 1981, Comb, 2008), and pY20 (Glenney et al., 1988). Overall, the three antibodies have a similar pY binding specificity whilst at the same time showing to some extent specific binding preferences in respect to tyrosine framing amino acids (Tinti et al., 2012). Nollau and Mayer (2001) performed a different approach to profile the global tyrosine phosphorylation state of cells by conducting one-dimensional PAGE and detecting tyrosine modifications of proteins by competitive binding to NRTK-specific recombinant SH2 domains by a far-Western approach.

Genetic screening and subsequent epistasis analysis conducted in several model organisms allowed successful identification of kinase substrates within signaling pathways (Manning and Cantley, 2002). For instance, perturbation of the cellular system by gene knock-outs or gene silencing by RNA inference and subsequent measurement of proteome-wide phosphorylation changes enabled identification of kinase-substrate and phosphatase-substrate relationships in *Drosophila melanogaster* (Friedman et al., 2011, Sopko et al., 2014).

Another interesting method is the creation of mutated kinases by genetic chemical design to track kinase-specific modifications inside the cell (Dephoure et al., 2005). In this “bump-and-hole” strategy, a large amino acid in the ATP binding site of the wild-type kinase is replaced by smaller amino acids creating a tiny “hole” widening the binding pocket. Thus, a bulky or “bumpy” radiolabeled or fluorescently labeled ATP analog can only be efficiently used by the exogenously expressed, mutated kinase whereby the wild-type kinase is removed from the kinome by gene knock-out or silencing. Attachment of labeled phosphate onto proteins can be tracked and therefore kinase-specific substrates identified (Dephoure et al., 2005).

Furthermore, flow cytometry was applied to determine protein modifications on a single cell level using phospho-specific antibodies conjugated to fluorophores (Perez and Nolan, 2002). Multiplexing is limited in this method by the spectral overlap of the employed flourophores. This limitation was overcome by combining flow cytometry with mass spectrometry creating a methodology termed mass cytometry (Bandura et al., 2009). Indeed, classical peptide-based mass spectrometry has

become a powerful technology for unbiased and hypothesis-free analysis of thousands of *in vivo* modifications with high accuracy and led to an dramatic increase in the number of annotated cellular phosphorylation sites (Olsen et al., 2006, Junger and Aebersold, 2014). Strategies for mass spectrometry (MS)-based proteomics are presented in the next section.

1.4.3. Mass spectrometry based determination of kinase substrates

The cellular proteome is defined by all proteins and their modifications at a given point in time. The generic workflow of MS-based proteomics is composed of four main procedures (Choudhary and Mann, 2010): 1) The proteome is extracted and subsequently fractioned or specific proteins or PTMs are enriched. 2) Proteins are digested into peptides which are separated and ionized. 3) Peptide and peptide fragment mass spectra are retrieved by mass analyzers. 4) Obtained data are analyzed by bioinformatics. Furthermore, MS-based analysis of peptides requires three types of information: (A) the peptide mass and (B) peptide fragments for identification, and the (C) signal intensity (e.g. spectral counts) for quantification (Choudhary and Mann, 2010). For each of the main steps in a generic MS workflow different methodologies have been developed with advantages and disadvantages strongly depending on the biological question to answer (e.g. depending of the PTM-of-interest). Often a combination of procedures is used maximize the output information. The workflow for a typical liquid chromatography tandem mass spectrometry (phospho-) proteomics (LC-MS/MS) experiment is presented in Figure 4.

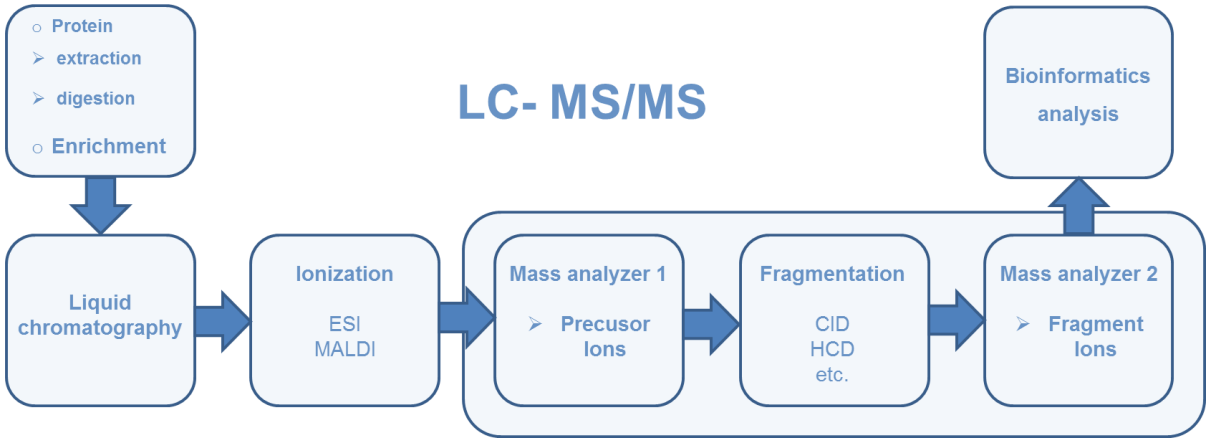


Figure 4: Typical workflow of liquid chromatography (LC) tandem mass spectrometry (MS/MS) in proteomics. Proteins are extracted from cell and either proteins-of-interest enriched and digested, or proteins digested and PTM-bearing peptides enriched. Peptides are separated by LC and ionized directly (electrospray ionization; ESI) or spotted (matrix-assisted laser desorption ionization; MALDI). Resulting Ions are passed through to the second mass analyzer (e.g. orbitrap), where they are measured at high resolution. Subsequently, precursor ions are selected in the first (ion trap) mass analyzer and subsequently fragmented by, for example, collision-induced dissociation (CID) or higher energy collisional dissociation (HCD) methods. Peptide fragments are measured in a second mass analyzer and peptides, peptide fragments, and PTMs identified using bioinformatics, i.e. by querying protein and peptide databases for the corresponding masses.

For the analysis of eukaryotic phospho-proteomes, it is not yet possible to achieve full coverage with current instrumentation and data analysis tools in single experiments (Junger and Aebersold, 2014). This is mainly due to magnitude differences in protein abundance. Furthermore, the high and increasing complexity of the phospho-peptidome compared to all unmodified peptides, the enormous number of measured and predicted phospho-sites, and low abundance of many signaling network components are not within the scale of the dynamic range of current LC-MS/MS instruments (Junger and Aebersold, 2014). In order to reduce the complexity of biological samples, and hence to increase the instruments' coverage, enrichment strategies for phospho-peptides are essential. Metal-based chromatography approaches employ binding of negatively charged amino acids to positively charged or polarized metal ions such as iron (Fe^{3+}) or titanium (Ti^{4+}) or to metal oxides such as titanium oxide (TiO_2) or zirconia oxide (ZrO_2) (Sopko and Andrews, 2008). Techniques termed immobilized metal-affinity chromatography (IMAC) and metal oxide affinity chromatography (MOAC) are commonly used to enrich for phospho-peptides from complex biological mixtures by immobilization of metal ions and metal oxides, respectively, on resins in a chromatographic column (Sopko and Andrews, 2008). Furthermore, small polar compounds can be separated on polar stationary phases by hydrophilic interaction liquid chromatography (HILIC) approaches as reviewed by Buszewski and Noga (2012) and the method was adapted to phospho-proteomics (Singer et al., 2010). Moreover, affinity purification using phosphorylation-specific antibodies or enrichment via chemical modification of phosphate groups using an affinity tag or a thiol group has also been used (Sopko and Andrews, 2008). In contrast to serine and threonine phosphorylation, there are highly specific antibodies recognizing tyrosine phosphorylation which perform well in physical isolation of tyrosine-phosphorylated macromolecules or peptides from a protease digest urea (Rush et al., 2005, Ballif et al., 2008, Junger and Aebersold, 2014). Antibody-based protein extraction has the advantage that it can be conducted under relatively harsh conditions such as 9M urea (Rush et al., 2005, Ballif et al., 2008, Junger and Aebersold, 2014). Furthermore, endogenous PKs were enriched using epitope or inhibitor displaying beads ("kinobeads") which can be used for the determination of kinase interactomes and potentially for identification of kinase-substrates via subsequent mass spectrometric analysis (Bantscheff et al., 2007). There are two general modes of peptide quantification: stable isotope labeling or label-free quantification. Label-free methods are based on the correlation between peptide abundance and the number of identified MS/MS spectra (spectral counts) or precursor ion signal intensity in combination with chromatographic information (Junger and Aebersold, 2014). Strategies employing isotope labeling may enable more accurate quantification of peptides (Olsen et al., 2006). These methods for peptide and protein quantification employ stable-isotope labeling of amino acids in cell culture (SILAC) (Oda et al., 1999), the use of isobaric tags chemically attached to peptides for relative and absolute quantitation (iTRAQ) (Ross et

al., 2004), or spiking labeled synthetic reference peptides into the biological samples to be quantified. Typically, proteins extracted or enriched from cell lysate are digested in gel or solution with protease Trypsin resulting in peptides carrying an arginine and lysine at the C-terminus. Separation of the proteome by 1D-PAGE and subsequent in-gel digestion is commonly followed by peptide separation via on-line liquid chromatography (LC) which is directly coupled to the Mass spectrometer. Peptides separated by LC are entering an ion source and are converted into intact ions in gas phase by electrospray ionization (ESI) or matrix-assisted laser desorption/ionization (MALDI) (Choudhary and Mann, 2010). Alternatively, in-solution digestion is succeeded by an additional peptide separation procedure such as ion exchange chromatography before on-line LC- tandem MS (Choudhary and Mann, 2010). There are different types of mass spectrometers as reviewed by Domon and Aebersold (2006) whereas in proteomics quadrupole time-of-flight (TOF) instruments and hybrid linear ion trap-orbitrap instruments are commonly used. In TOF instruments, peptide ions are separated by the mass and charge-dependent time they require to pass through a chamber until arrival at a detector. In contrast, orbitrap instruments measure the frequency of peptide ions oscillating inside the ion trap dependent only on their mass and charge. Mass spectra are retrieved by converting the frequency signal by a mathematical operation namely Fourier transformation. Of note, an important parameter is the mass spectrometric resolution, a unit-less quantity to define how well peptide ions can be distinguished at the time of co-elution from the on-line chromatographic column. It is approximately 20000 for TOF instruments and five to six times higher for orbitrap instruments allowing analysis of samples of greater complexity (Choudhary and Mann, 2010, Newman et al., 2014). A major challenge in phospho-peptide detection is the instability of the phosphate monoester linkage which will be cleaved to some extent by bombardment with inert gases in collision-induced dissociation (CID) methods. Hence, “phosphate-friendly” fragmentation methods are employed such as higher energy collisional dissociation (HCD) (Newman et al., 2014). Generally, modern MS-instruments can capture almost the entire phospho-proteome of an organism. However, targeted MS-approaches employing single reaction monitoring (SRM) or multiple reaction monitoring (MRM) are more applicable for quantification and verification of specific, known phospho-peptides in complex mixtures (Junger and Aebersold, 2014). Upon detection in the mass analyzer sequences of peptide ions and peptide fragment ions are stored and mapped to the proteome-of-interest by search algorithms including SEQUEST (Eng et al., 1994), MASCOT (Perkins et al., 1999), X!Tandem (Craig and Beavis, 2004), the OpenMS Search Algorithm (Sturm et al., 2008), and Andromeda (Cox et al., 2011) - the latter as part of the MaxQuant environment (Cox and Mann, 2008). Phosphorylation sites from low- and high-throughput studies are stored in publicly accessible databases such as PhosphoSitePlus (phosphosite.org) (Hornbeck et al., 2012), SwissProt (Farriol-Mathis et al., 2004), HPRD (Prasad et al., 2009), and Phospho.ELM (Dinkel et al., 2011).

Examples of MS-based *in vivo* phospho-proteomics studies employing different enrichment strategies are presented in the following paragraph. The combination of ion exchange chromatography with nanoflow high performance liquid chromatography (HPLC)/electrospray ionization mass spectrometry was the successfully applied to measure the phospho-proteome of yeast and of capacitated human sperm cells using whole-cell lysate (Ficarro et al., 2002, Ficarro et al., 2003). In their approach Ficarro and colleagues converted acidic residues to methyl esters to block their binding to iron during IMAC and thus prevents contamination of the column with non-phosphorylated peptides. By their approach dozens (human sperm) or hundreds (yeast) of phospho-peptides were identified however, only a very small fraction of identified peptides contained phospho-tyrosine (pY). Instead of employing IMAC, Rush et al. (2005) used a pY-specific antibody to enrich for pY-peptides together with LC-MS/MS analysis and was hence able to detect 628 mainly novel tyrosine-phosphorylation sites (pY-sites) from cancer cell lysates. Immuno-affinity purification using phosphoamino acid-specific antibodies was employed in later studies collecting phospho-sites from various cells under various conditions (Rikova et al., 2007, Ballif et al., 2008, Kruger et al., 2008, Boersema et al., 2010). For instance, Rikova et al. (2007) performed tyrosine phosphorylation profiling on 41 non-small cell lung cancer (NSCLC) cell lines and over 150 NSCLC patient-derived tumors using the antibody-based strategy of Rush et al. (2005). Thus, the authors could identify 4551 tyrosine phosphorylation events on more than 2700 different proteins whereas 85 percent of the sites appeared novel. Thus, phospho-tyrosine signaling was comprehensively compared across many different NSCLC samples which were grouped and characterized according to the activation status of identified tyrosine kinases (Rikova et al., 2007).

Both types of MS-based approaches, employing either affinity columns, or immuno-affinity phospho-peptide enrichment led to the identification of modified tyrosine residues however did not yield any information about responsible PTKs. In order to link kinases to substrates, Matsuoka et al. (2007) used kinase-specific phosphorylation site recognizing antibodies for peptide enrichment and hence was able to determine 900 phosphorylation sites regulated by ATM (ataxia telangiectasia mutated) and ATR (ATM and Rad3-related) kinases encompassing over 700 proteins. ATM and ATR are kinase activated upon DNA damage and resemble the core of the DNA damage signaling, and share substrate specificity for Ser-Gln (SQ) and Thr-Gln (TQ) motifs (Matsuoka et al., 2007). 293T cells irradiated with infra-red light were compared to untreated cells by SILAC whereat putative ATM and ATR substrates identified by phospho-immuno-affinity enrichment using antibodies to phospho-SQ or phospho-TQ sites (Matsuoka et al., 2007). To identify substrates specific for MEK1 kinase, Morandell et al. (2010) used ATP-depleted cytosolic extracts of MEK1 deficient MEFs for *in vitro* kinase assays on the principles of chemical genetics. An ATP analog sensitive MEK1 mutant was expressed in the

presence of “bumpy” ATP and after cell lysis phospho-peptides were enriched using IMAC followed by identification via tandem MS. The direct comparison to cell lysate treated with normal ATP lead to identification of two known MEK1 substrates ERK1 and ERK2 however, not a single novel target. Holt et al. (2009) determined Cdk1 substrates in yeast by measuring the loss of phosphorylation upon chemical Cdk1 inhibition via quantitative mass spectrometry. In their SILAC-based approach, isotope-labeled proteins were extracted from cells, digested, resulting peptides were separated into 12 fractions by strong cation exchange (SCX) chromatography, and for each fraction phospho-peptides were enriched via IMAC (and TiO_2). In this way, 547 phosphorylation sites on 308 Cdk1 *in vivo* substrates were discovered (Holt et al., 2009). In order to identify substrates of yeast DNA damage checkpoint kinases, Smolka et al. (2007) performed *in vivo* determination of the yeast phospho-proteome upon DNA-damage comparing kinase knock-out cells to wild-type cells. After treatment with the DNA alkylating agent methyl methane sulfonate (MMS) cells were lysed, proteins extracted and Trypsin digested, and phospho-peptides enriched by IMAC and labeled with an amine-reactive isotope tag for quantitative LC-MS/MS analysis. Among 2689 phosphorylation sites measured they observed a loss of 62 phosphorylation sites from 55 proteins upon specific knock-out of key enzymes which were in part verified as direct targets by *in vitro* kinase assay. It is noteworthy that many candidate substrate phosphorylation sites may be missed by this approach as other unknown kinases may act on the same sites preventing the observation of a loss upon kinase knock-out. A system-wide, comprehensive screen for yeast kinase and phosphatase substrates was conducted by Bodenmiller et al. (2010). The authors created 116 kinase-depleted or phosphatase-depleted mutant yeast strains and eight strains expressing mutants which are inhibitable by cell-permeable drugs. The phospho-proteome of each knock-out strain was measured by LC MS/MS using an enhanced protocol of TiO_2 chromatography for phospho-peptide enrichments (Larsen et al., 2005) and compared to wild-type strains. Even though a large fraction of the kinase and phosphatase regulated 8814 phosphorylation events was considered to be indirect, Bodenmiller et al. (2010) could generate the first system-wide phosphorylation network describing the relationships between 97 kinases, 27 phosphatases, and more than 1000 substrates. All methods aiming to identify responsible kinase suffer limitations in defining *bona fide* KSRs due to the complexity of kinase action in its native environment. Apart from indirect determination of substrates by measuring the loss of phosphorylation sites upon elimination of the kinase, responsible enzymes are usually inferred by querying measured phosphorylation site sets for kinase specificity determinants such as linear sequence motifs. An overview of methods developed to define kinase specificity for ultimately linking kinases to the vast amount of measured phosphorylation sites is provided in the next section.

1.4.4. Determination of PTK specificity

Due to the very large number of potential substrates and phosphorylation sites presented to a kinase in any cellular environment, kinases must ensure to some extent precise substrate targeting in order to maintain signaling fidelity. Kinase specificity determinants provided by the cellular environment include subcellular localization, substrate docking interactions, complex formation with scaffolding and adaptor molecules, and systems-level competition between substrates (Ubersax and Ferrell, 2007). Besides these more distant specificities, the local substrate protein structure and primary amino acid sequence surrounding the phosphorylation site regulate binding of the kinase catalytic domain and constitute major factors in substrate recognition (Miller et al., 2008, Duarte et al., 2014). Recurrences of amino acids flanking kinase specific target sites can be represented as linear amino acid sequence motifs which can be employed to predict kinase substrates (Miller et al., 2008, Mok et al., 2010, Newman et al., 2013). Due to the limited number of known kinase substrates direct measurement of amino acid frequencies is often infeasible. Thus, oriented peptide libraries were used to determine the optimal binding sequence for individual kinase domains (Songyang et al., 1994, Hutti et al., 2004, Miller et al., 2008, Mok et al., 2010). In these approaches, a positional scanning peptide library is generated consisting of peptide mixtures. In each mixture one of the 20 naturally occurring amino acids is fixed at each of seven to nine positions surrounding a central non-phosphorylated or phosphorylated (for comparison) serine, threonine, or tyrosine residue, with the remaining positions degenerate. The resulting peptide library is incubated in parallel with individual kinases and radiolabeled ATP for phosphorylation reactions. Incorporation of radiolabeled phosphate into each of the peptides is visualized and quantified by transfer of the reactions to an avidin-coated membrane. Substrate peptides bind to the membrane by a C-terminal biotin tag and remain bound during extensive washing steps (Hutti et al., 2004). There are different types of motifs with differing levels of information content and complexity. Short consensus phosphorylation motifs such as S/T-P-X-R/K for CDK1 (Alexander et al., 2011) where residues most frequently occurring at particular positions relative to the modification site are indicated (such as proline at one position C-terminal to phosphorylated serine or threonine in the provided example), are employed by the algorithms of Scansite (Obenauer et al., 2003) and Minimotif Miner (Mi et al., 2012). These simple representations are easy to understand and easy to use however, may in general not be sufficient to describe specificity for a kinase substrate (Joughin et al., 2012). A more complex representation of sequence specificity comprise linear motifs based on position-specific scoring matrices (PSSM) which indicate the probability of residues to occur around a modification site based on oriented peptide library screening (Hutti et al., 2004), for instance. Both of representations rely solely on local primary sequence information surrounding potential phosphorylation sites however, they differ in the determination of the suitability of potential substrates (Joughin et al., 2012). In contrast to short

linear motifs where only optimal substrates sequences matching the most frequently co-occurring residues may be identified, PSSM-based motifs can also capture suboptimal substrates. When using short linear motifs, substrates are identified by a Boolean descriptor ("YES" and "NO") under the assumption a kinase is targeting the substrate sequence when it fits the motif and does not target the substrate if the modification site does not match exactly the motif sequence. On the contrary, probability based linear motifs provide a score via the PSSM for each putative substrate sequence dependent on how well the modification site sequence matches the motif of a putative kinase. Finally, more complex models of kinase specificity descriptors than PSSM-based motifs were developed employing machine-learning algorithms. These algorithms use artificial neural networks such as NetPhosK (Blom et al., 2004) and NetPhorest (Miller et al., 2008), hidden Markov models as implemented in KinasePhos (Huang et al., 2005), Bayesian decision theory as in the program Prediction of PK-Specific Phosphorylation site (PPSP) (Xue et al., 2006), or support vector machines (Kim et al., 2004). Another algorithm called Group-based Prediction System (GPS) (Xue et al., 2008) infers kinase substrates by structural and sequence similarity. GPS employs amino acid substitution matrices in order to improve sequence based-classifiers by integration of higher-order information. The strength of theses learning algorithms to improve substrate identification highly depends on the size of the training sets, i.e. the number of experimentally identified kinase substrates stored in databases such as PhosphoSitePlus (phosphosite.org) (Hornbeck et al., 2012), SwissProt (Farriol-Mathis et al., 2004), HPRD (Prasad et al., 2009), and Phospho.ELM (Dinkel et al., 2011) as mentioned earlier. These complex models potentially fail due to heterogeneity within in the necessary large databases and resulting training sets (Joughin et al., 2012). Database heterogeneity emerged by collecting substrates identified by different experimental approaches under different conditions which may be biased (Joughin et al., 2012). Importantly, addition of contextual information to improve motif-based kinase-substrate predictions is valuable as primary sequence information alone may explain only approximately 20 percent of the overall kinase specificity (Linding et al., 2007). Thus, integrative *in-silico* workflows were developed including contextual information by integrating PPI networks as in NetworkKIN (Linding et al., 2007) and PhoshoPICK (Patrick et al., 2015) for building kinase-substrate predictors. There are several motif extraction tools integrated in workflows to query sequence data sets including WebLogo (Crooks et al., 2004), Motif-X (Schwartz and Gygi, 2005) and Scan-X (Schwartz et al., 2009), PhosphoMotif Finder (Amanchy et al., 2007), MoDL (Ritz et al., 2009), pLogo (O'Shea et al., 2013), and iceLogo (Colaert et al., 2009).

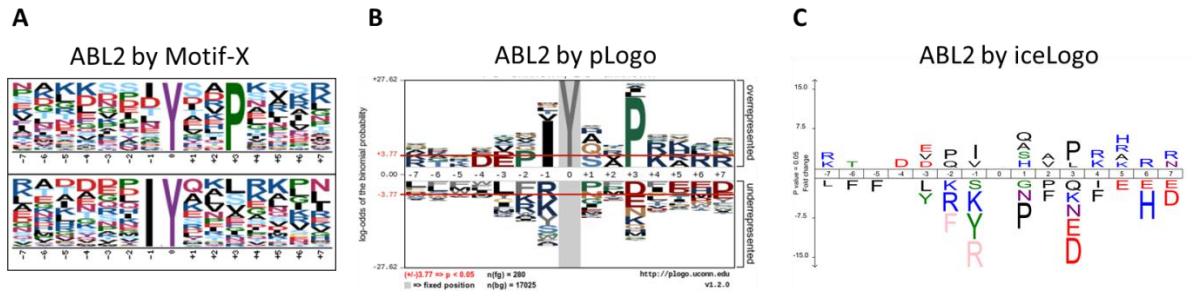


Figure 5: Linear sequence motifs exemplified using 280 ABL2 yeast target sites (A) Frequency logos extracted with motif-x (Schwartz and Gygi, 2005). The upper motif represents 62 of 280 pY-sites, the motif below represents 44 of (280-62=) 218 pY-sites. (B) Enrichment motif generated with pLogo (O'Shea et al., 2013). (C) Enrichment motif obtained with iceLogo (Colaert et al., 2009).

There two main types of sequence motif representations. On the one hand, there are frequency plots such as those obtained by Motif-X where the frequency of amino acids within a subset of substrate sequences is presented (Figure 5A). By an iterative approach, motif-x queries input sequences for overrepresented residues amino acids above a probability cut-off in comparison to a background sequence set. The obtained sequences are subtracted from the original set and the remaining sequences are re-analyzed for further enrichment of amino acids within the reduced set. The two frequency plots for ABL2 targets given as an example output of motif-x are thus containing together only 66 pY-sites of 280 (Figure 5A). On the other hand, there are enrichment logos as obtained by pLogo and iceLogo showing the amino acid frequencies within substrate sequence sets compared to genomic background, i.e. the frequency of enriched amino acids relative to the frequency of suchlike encoded by the genome (Figure 5B and C). Both pLogo and iceLogo are probability-based however, differ from each other in the metric by which individual residues are scaled and in the background model assumed (O'Shea et al., 2013). A feature provided only by pLogo is the possibility to fix the position of another residue to determine inter-positional dependencies. The two enrichment plots contain all 280 ABL2 yeast target sequences and both show amino acid over- and underrepresentation compared to genomic background. While both types of motifs are informative, probabilistic enrichment motifs may comprise higher information content enabling prediction of suboptimal substrates as opposed to the frequency plots which are rather similar to short linear motifs.

The performance of motifs in defining kinase targeting specificity strongly depends on the number of sequences provided. Thus, combinatorial linear peptide libraries and functional protein microarrays were employed for determination of *in vitro* kinase substrates and the optimal linear amino acid sequence motifs (Miller et al., 2008, Mok et al., 2010, Newman et al., 2013). Linear peptide libraries overcome limitations in the availability of known kinase substrates. Generated optimal consensus

linear sequence motifs are integrated into bioinformatics tools for kinase substrate predictions in combination with known kinase-specific phosphorylation sites (Newman et al., 2013) and contextual information such as protein-protein interaction (PPI) networks by NetworKIN (Linding et al., 2007) and kinomeXplorer (Horn et al., 2014). Moreover, data integration using phosphorylation site, motif and signaling pathway information together with gene ontology (GO) term enrichment analysis was performed to generate kinase/phosphatase/substrate signaling networks (Petsalaki et al., 2015). In order to retrieve not the optimal, but actually targeted linear sequences researchers try to mimic the expressed human proteome as basis for kinase phosphorylation reactions. Thus, human cell lysate was treated with proteolytic enzymes prior to incubation with enzymes-of-interest providing a database-searchable proteome-derived peptide library in order to identify endoprotease cleavage sites (Schilling and Overall, 2008), for instance. Those proteome-derived peptide libraries from trypsin treatment were dephosphorylated using phosphatases to subsequently perform *in vitro* kinase assays for defining kinase specificity and substrates as conducted by Wang et al. (2013), for example. In order to improve the sensitivity and specificity of the *in vitro* kinase reaction Xue et al. (2012) additionally identified in parallel *in vivo* kinase substrates by quantification of *in vivo* phosphorylation sites by the endogenous kinase compared to kinase knock-out cells. The combination of the two datasets may increase confidence in identification of putative kinase substrates (Xue et al., 2012). During preparation of this theses Chou et al. (2012) published a strategy dubbed ProPeL (Proteomic Peptide Library) in which they exogenously expressed serine/threonine (S/T) kinases in intact bacteria and measured phosphorylation events within the bacterial proteome relying on the assumption that human S/T kinases are sufficiently different to bacterial S/T kinases. The authors could show that motifs retrieved from human kinase activity on the bacterial proteome performed better in substrate prediction than motifs generated based on human proteome-derived peptide libraries linearized and de-phosphorylated after cell lysis (Chou et al., 2012).

1.5. Thesis aims

Tyrosine protein phosphorylation in human is highly abundant and plays a major role in signal transduction, cell to cell communication, cellular growth, and proliferation. Deregulation of tyrosine phosphorylation is observed in many complex human diseases including cancer. However, defining KSRs conceptually and experimentally remains a challenging task even though many different experimental and computational approaches have been developed since the discovery of SRC kinase in the late eighties. This thesis describes a conceptually novel approach of determining human NRTK specificity using the lower eukaryote *Saccharomyces cerevisiae* as model organism. Tyrosine signaling can be regarded a hallmark of multi-cellularity and has not evolved in yeast (Manning et al., 2002a). Upon weak exogenous expression of individual full-length human cytoplasmic PTKs in yeast, tyrosine phosphorylation of fully-folded, intact yeast proteins can be measured employing antibody-based phospho-tyrosine peptide enrichment followed by mass spectrometry analysis (Rush et al., 2005). This phosphorylation events can be unambiguously attributed to specific human kinases in a background-free manner. In this way, yeast protein targets for 16 out of 32 NRTKs representing at least one member of almost all NRTK families (Blume-Jensen and Hunter, 2001) could be measured comprehensively in a single, cellular set-up. The obtained tyrosine phosphorylation data enabled inter-NRTK substrate comparisons. Furthermore, the method allowed to determine human orthologous pY-sites and inference of responsible NRTKs. Thus, targeting of glycolytic enzymes by NRTKs was discovered linking phospho-tyrosine signaling with cellular metabolism. Primary sequence specificities were obtained from the yeast data in form of well-performing linear sequence motifs which enabled inference of NRTKs for 1388 human pY-sites – approximately ten percent of all by the time reported 13240 human pY-sites. Motif-based predictions suggested NRTK targeting of single substrates and substrate families such as CDKs. Finally, direct targeting of the yeast protein-protein interaction network was observed and characterized. In summary, a large and unique dataset was generated by a novel strategy enabling determination of targeting specificity of individual human cytoplasmic PTKs in a cellular environment and subsequent assignment of NRTKs to reported human tyrosine phosphorylation sites.

2. Material and methods

2.1. General molecular biology methods

2.1.1. Gateway cloning

Gateway cloning is an efficient cloning technique developed by Invitrogen Inc. based on the recombination system of phage lambda. DNA sequences can be easily shuttled between vectors employing a proprietary set of recombination sequences and two proprietary enzyme mixes. The recombination sequences or attachment sequences ("att sites") are attached to the ends of DNA strands of interest (e.g. ORFs) via PCR and appropriate primers. This enables recombination and insertion of the DNA sequence in plasmids using two enzyme mixes "BP-clonase" and "LR-clonase". In contrast to conventional cloning strategies, gateway cloning can yield up to 99 percent positive transformants.

2.1.1.1. BP-reaction

Any DNA sequence with specific gateway recombination sites (here *attB* sites) attached can be combined with gateway "donor vectors" pDONR221 or pDONR223 (containing *attP* sites) and the "BP-clonase" enzyme mix yielding a gateway "entry vector" or "entry clone". This entry clone containing *attL* recombination sites can be used to shuttle the DNA into various kinds of compatible expression vectors. Usually, 75 ng entry vector, 20 ng PCR product and 1 µl BP clonase enzyme mix II was mixed on ice and the final volume adjusted to 5 µl using TE buffer pH 8.0. After mixing by vortexing, the reaction was carried out over-night at 25°C in a polymerase chain reaction (PCR) machine. In order to stop the reaction 0.5 µl Proteinase K were added and enzymes digested at 37°C for 15 min. It is possible to store the reaction product at -20°C. Correct insertion of the DNA was analyzed by agarose gel electrophoresis after plasmid transformation into *E. Coli*, plasmid extraction, and digestion using restriction enzyme BsrG1 (see below).

2.1.1.2. LR-reaction

Generated Entry clones contain an ORF or gene of interest fused to *attL* sites which can be rapidly shuttled into various yeast expression vectors (gateway "destination vectors") harboring the appropriate recombination sites (*attR* sites). After mixing 1 µl "LR-clonase mix II" with 75 ng destination vector and 200-300 ng entry vector the total volume was adjusted to 5 µl using TE buffer pH 8.0. The reaction was carried out over-night at 25°C in a PCR machine. In order to stop the reaction 0.5 µl Proteinase K were added and enzymes digested at 37°C for 15 min. It is possible to store the reaction product at -20°C. Correct insertion of the DNA insert, harboring *attB* sites again,

was analyzed by agarose gel electrophoresis after direct transformation of the reaction into *E. Coli*, plasmid extraction, and digestion using the restriction enzyme BsrG1 (see below).

2.1.2. Polymerase Chain Reaction

For polymerase chain reaction (PCR) different enzymes were used initially such as high-fidelity Phusion DNA polymerase (Stratagene, Santa Clara, USA) and proof-reading PfuTurbo DNA Polymerase (Stratagene, Santa Clara, USA). However, for the majority of performed PCR reactions high-fidelity, proof-reading KOD DNA polymerase (Novagen/Toyoba, Merck Millipore Corporation, Billerica, MA, USA) was used which combines the advantages of the latter two.

KOD PCR reaction mixture

10 ng template

10 µM each primer (fwd + rev)

0.5 µl KOD polymerase (1U/µl)

2.5 µl 10x KOD PCR buffer

2.5 µl dNTP 2mM each

1,5 µl $MgSO_4$ 25mM

1,5 µl DMSO (PCR grade)

H_20 to a total volume of 25 µl

PCR reaction protocol

1. 2 min 95°C
2. 20 sec 95°C
3. 10 sec 55°C to 65°C
4. 22 sec 70°C

Repeat 24 times step 2 to 4

5. 2 min 70°C
6. 4°C for ever

2.1.3. Determination of DNA concentration

DNA concentrations in was determined aqueous solution by measuring the absorbance at 260 nm in a spectrophotometer (NanoDrop ND-1000). 1µl of DNA solution is applied to the instrument and absorbance measured against a blank sample (buffer or H_2O only).

2.1.4. DNA sequencing

Automated fluorescent DNA sequencing was performed by the in house sequencing service (MPI for Molecular Genetics) or StarSEQ® GmbH (Mainz, Germany) sequencing service to verify ORF identity within gateway entry or gateway destination vectors. Primers used for sequencing were synthesized as described below.

Symbol	Sequence	Direction	Vector
M13_f	TGTAAAACGACGGCCAGT	fwd	pDONR
T7_r	TAATACGACTCACTATAGGG	rev	pDONR
CUPSeq-5p	CTTGTCTTGTATCAATTGC	fwd	pASZ_C/CN
CycT-5m	GGACCTAGACTTCAGGTTG	rev	pASZ_C/CN

Table 1: Primer used for sanger sequencing of clones.

2.1.5. Restriction digest

The restriction endonuclease Bsp1407I, a fast digest isoschizomer of BsrGI, was used to control for correct ORF length within plasmids originating form gateway reactions via agarose gel electrophoresis. Typically, 5 µl DNA sample were mixed with 15 µl restriction endonuclease mix (2 µl 10x FastDigest Buffer, 0.1 µl FastDigest Bsp1407I, 12.9 µl H_2O) and incubated at 37 °C for 45 minutes. Heat inactivation at 65°C for 10 minutes was performed.

2.1.6. Agarose Gel-electrophoresis

2.1.6.1. Buffers and Reagents

10x Tris/Borate/EDTA (TBE) buffer

108 g Tris Base

55 g Boric acid

40 ml 0.5 M EDTA pH 8.0

pH 8.3 (using HCl)

H_20 to a total volume of 1 liter

10x Orange G sample buffer

50 % (w/v) Sucrose

0.5 % (w/v) Orange G

2.1.6.2. Gel-electrophoresis and gel staining

Linearized DNA was analyzed via agarose gel-electrophoresis. Therefore, 0.8 to 1.2 % agarose gels were prepared by completely dissolving 0.8 to 1.2 (w/v) electrophoresis grade agarose in 0.5 % TBE buffer in a microwave and, after cooling down, pouring the solution into a gel chamber. DNA was mixed with Orange G sample buffer and loaded onto the agarose gel. Additionally, 7 µl of a DNA

marker (1 Kb Plus DNA ladder from Invitrogen) was loaded in one gel lane for DNA band size estimation. Gels ran on differing voltage and time dependent on gel and chamber size and percentage of agarose in 0.5 % TBE running buffer. For example, a one percent 50 ml gel ran for 40 min at 75 volts. After running, the gels were stained for 20 min in a 1:20000 dilution of SYBR Gold nucleic acid gel stain in 0.5 % TBE and analyzed by UV light excitation in a gel documentation apparatus from Intas Science Imaging Instruments GmbH.

2.1.7. SDS Gel-electrophoresis

2.1.7.1. Reagents and Solutions

Separating gel 12% (80ml)

25 ml 40% AA/BAA (37.5:1)

20 ml 1.5 M Tris pH 8.8

800 µl 10 % SDS solution

35.2 ml H_20

60 µl TEMED

600 µl Ammonium persulfate (APS)

Stacking gel (40ml)

5 ml 40% AA/BAA (37.5:1)

20 ml 1.5 M Tris pH 8.8

800 µl 10 % SDS solution

29.6 ml H_20

80 µl TEMED

400 µl Ammonium persulfate (APS)

4x SDS gel loading buffer

200 mM Tris-Cl (pH 6,8)

4 % SDS

40 % glycerol

0,4 % bromphenol blue

Added prior to use:

200 mM DTT

2.5x alternative SDS gel loading buffer

200 mM Tris-Cl (pH 6,8)

5 % SDS

25 % glycerol

25 mg bromphenol blue

Added prior to use:

12.5% v/v β-mercaptoethanol

10x SDS gel electrophoresis buffer

30,2 g Tris Base

144 g glycine

10 g SDS

H_20 to a total volume of 1 liter

2.1.7.2. Method

12 % SDS gels were produced using a gel chamber for making 12 gels at once or single gels using a Biorad system. All glass equipment was cleaned with 70 % Ethanol and gel chambers assembled. The separating gel solution was prepared and filled into the chamber. On top of the separation gel, a thin layer of isopropanol was pipetted to prevent the gel from drying out during polymerization for one hour at room-temperature. As a next step, the stacking gel solution was prepared and immediately added on top of the separation gel. For creating loading pockets plastic combs were inserted immediately after pouring the gel solution and the gel was allowed to polymerize for one hour at room-temperature. Protein samples, e.g. cell lysate, was mixed with SDS loading buffer and heated to 95°C for 5 min. SDS gels were placed into the electrophoresis chamber according to the manufacturer's protocol. The protein samples (5 to 15 µl) were loaded into the gel pockets and the gel ran 15 min at 90V to collect the proteins in the stacking gel. Subsequently, the gel ran for 1 hour at 120V until the SDS loading buffer left the separation gel indicating the smallest proteins arrived at the end of the gel.

2.1.8. Western-blotting

2.1.8.1. Reagents and Solutions

10x Tris-buffered saline (TBS)

30 g Tris Base

87 g NaCl

H_20 to a total volume of 1 liter

1x TBST membrane wash solution

10 % v/v 10x Tris-buffered saline (TBS)

0.1 % Tween 20

H_20 until total volume reached

Membrane blocking solution

5 % w/v bovine serum albumin (BSA)

OR 5 % w/v non-fat milk powder

1x TBST (0.1% Tween 20)

Antibody solutions

3 % w/v bovine serum albumin (BSA)

antibody in appropriate dilution

1x TBST (0.1% Tween 20)

10x wet-blot transfer buffer without methanol

30,2 g Tris Base

144 g glycine

H_20 to a total volume of 1 liter

1x wet-blot transfer buffer

100 ml 10x Wet-blot transfer buffer

200 ml 100% methanol

700 ml H_20 to a total volume of 1 liter

10x semi-dry transfer buffer without methanol

58,1 g Tris Base

29,3 g glycine

3,75 g SDS

H_20 to a total volume of 1 liter

1x semi-dry-blot transfer buffer

100 ml 10x Wet-blot transfer buffer

200 ml 100% methanol

700 ml H_20 to a total volume of 1 liter

2.1.8.2. Blotting Methods

For wet-blotting a wet-blot transfer buffer was prepared freshly. The SDS gels were placed into the blotting chamber according to the manufacturers protocol. In brief, inside the holder on top of a thick chromatography paper the nitrocellulose or PDVF membrane (which has to be activated in 100 % methanol) was placed followed by the SDS-gel and one additional chromatography paper. All layers were placed within the pre-cooled transfer buffer, completely covered, in order to avoid air-bubbles. The sandwich was subsequently placed vertically in the blotting chamber so that the cathode is at the side of the gel and the proteins are hence pulled out of the gel onto the membrane. The blotting was performed in the cold-room at 4°C with an ice-pack added into to chamber for compensating heat generation during blotting. A blot typically ran for 70 min at 120 V.

For semi-dry blotting the sandwich as described above was assembled on top of the horizontal cathode in the semi-dry chamber and covered in transfer buffer. Importantly, air-bubbles have to be avoided by applying pressure on the sandwich with a roller (e.g. a cut 25 ml plastic pipette). The blot ran with adjusted current (0.8 mA per cm^2 of membrane/gel) for 1 hour at room-temperature. The advantage towards wet-blotting is the use of less buffer and the higher running capacity per chamber (4 gels instead of 2 gels per chamber for the wet-blot) however, the disadvantage is that high molecular weight bands are less efficiently or not at all transferred.

2.1.9. Protein staining

2.1.9.1. Solutions

Coomassie-Blue stain

45% v/v methanol (100%)

10% v/v glacial acetic acid

45% v/v H_20

3 g/L Coomassie Brilliant Blue R250

Coomassie-Blue Destain solution

45% v/v methanol (100%)

10% v/v glacial acetic acid

45% v/v H_20

Blue silver stain

20% v/v methanol (100%)

10% v/v phosphoric acid

10% v/v ammonium sulfate

3 g/L Coomassie Brilliant Blue G-250

60% v/v H_20s

Ponceau membrane stain

0,5% w/v Ponceau S

1% v/v glacial acetic acid

H_20 to a total volume of 1 liter

2.1.9.2. Methods

After SDS- gel electrophoresis proteins were visualized within the gel using Coomassie-Blue stain or the more sensitive blue silver stain. Gels were placed into the solution and incubated for 20 min at room-temperature on a shaker. Subsequently, gels were de-stained for 20 min or up to several hours in Destain-solution (staining solution without dye) at room-temperature until only the protein bands remained stained. Pictures were taken on a light table. Gels were be stored at 4°C in de-stain solution or water. Proteins can be also visualized less quantitatively on a blotting membrane. Therefore, nitrocellulose or PDVF membranes were placed in Ponceau solution and incubated 5 min at room-temperature on a shaker. Membranes were de-stained by rinsing with pure water and pictures taken on a light table.

2.2. Escherichia coli (*E. coli*)

2.2.1. *E. coli* strains

DH10B: F^-, mcrA, Δ(mrr hsd RMS-mcr BC), φ80dlacZΔM15, ΔlacX74, deoR, recA1, araD139, Δ(ara, leu)7697, alU, galK, λ^-, rpsL, endA1, nupG (Invitrogen)

DB3.1: F^-, gyrA462, $endA^-$, Δ(sr1-recA), mcrB, mrr, hsdS20(r_B^- m_B^-), supE44, ara14, galK2, lacY1, proA2, rpsL20(Sm^r), xyl, λ^-, leu, mtl1

2.2.2. *E. coli* growth media

LB-medium

10 g NaCl

5 g Bacto yeast extract

10 g Bacto tryptone

pH 7.2 (adjusted with NaOH)

H_20 to a total volume of 1 liter

LB-agar

10 g NaCl

5 g Bacto yeast extract

10 g Bacto tryptone

pH 7.2 (adjusted with NaOH)

20 g Bacto agar

H_20 to a total volume of 1 liter

2YT-medium

5 g NaCl

10 g Bacto yeast extract

16 g Bacto tryptone

pH 7.2 (with NaOH)

H_20 to a total volume of 1 liter

Transformation and storage solution (TSS)

85 % LB-medium

10 % (w/v) PEG 8000

5 % DMSO

50 mM $MgCl_2$

Filter sterilize through a 0.22 µm pore filter

SOB-medium

0.5 g NaCl

5 g Bacto yeast extract

20 g Bacto tryptone

H_20 to a total volume of 1 liter

after autoclaving:

10 ml 1 M $MgCl_2$

10 ml 1 M $MgSO_4$

SOC-Medium

99 ml SOB-medium

1 ml 20x Glucose

Antibiotic	concentration
Ampicillin	100 µg/ml
Chloramphenicol	30 µg/ml
Tetracycline	20 µg/ml
Kanamyci	15 µg/ml
Spectinomycin	50 µg/ml

2.2.3. Solutions for plasmid isolation of bacteria

Buffer P1	*Buffer P2*	*Buffer P3*
50 mM Tris pH 8.0	200 mM NaOH	3 M Potassium acetate pH 5.5
10 mM EDTA pH 8.0	1 % (w/v) SDS	1 M Glacial acetic acid
50 mg/l RNase A		

2.2.4. Generation of competent *E. coli*

2.2.4.1. Chemically competent *E. coli*

Bacteria were rendered chemically competent for taking up nucleic acids (i.e. plasmids) via heat-shock. The bacteria used for this procedure were DH10B whereas a freshly grown, single colony was picked into 20 ml 2YT-medium and incubated at 37°C on a shaker over-night (up to 20 hours). The overnight culture was used to inoculate 2 liter 2YT-medium and incubated shaking at 37°C until an optical density of 0.7 to 0.8 was measured at a wavelength of 600 nm (OD_{600}). For subsequent centrifugation at 2250 x g for 15 minutes at 4°C the culture was divided in four. The supernatant was discarded and each pellet re-suspended in 40ml sterile Transformation and Storage Solution (TSS).For storage at -80°C 4 ml 87 % sterile glycerol was added prior to freezing.

2.2.4.2. Electroporation competent *E. coli*

Bacteria need to be competent for taking up nucleic acids (i.e. plasmids) via electroporation. In order to generate electroporation competent DH10B a freshly grown, single colony was picked into 50 ml LB-medium and incubated over-night shaking at 37°C. Having grown 16-20 hours 45 ml of the over-night culture was transferred to 1.5 liter LB-medium and incubated shaking at 37°C until an optical density of 0.3 to 0.4 was measured at a wavelength of 600 nm (OD_{600}). For subsequent centrifugation

the culture was divided into six aliquots and placed on ice for 30 minutes. After subsequent centrifugation at 3500 rpm for 10 minutes at 4 °C the supernatant was discarded and each pellet was re-suspended in 100 ml cold, sterile 15 % glycerol. The procedure was repeated twice with 20 minutes placement on ice and successive decreasing volumes to 10 ml and 2 ml re-suspension in cold, sterile 15 % glycerol. Finally, aliquots of the cells were and stored at -80°C

2.2.5. Transformation of competent *E. coli*

2.2.5.1. Chemical transformation

The procedure presented here is a high-throughput version of the standard procedure using a 96 well plate format for transforming plasmid products of the gateway cloning system ("BP" and "LR" reactions) into bacteria. As a first step, chemically competent cells were thawed on ice and 30 µl aliquoted into a 96 well PCR plate. To each well 1 µl to 3 µl of BP or LR reaction was added and the plate gently vortexed, sealed with tape and incubated on ice for 30 min. The cells were transformed using a 90 second heat-shock at 42°C and subsequently incubated on ice for 5 min. The entire content of each well was transferred to the corresponding well of a 96 deep-well plate pre-filled with 270 µl of 37°C warm SOC medium. The deep-well plate was incubated for 1 hour at 37°C while shaking to allow the cells to recover. Finally, 10 µl of each well was spotted multiple times (10-12 spots) in rows on selective LB-agar plates and the incubated over-night at 37°C. If the transformation efficiency was too low, an electroporation protocol was applied instead.

2.2.5.2. Electroporation

Electroporation competent cells were thaw on ice and 50 µl aliquots were transferred into micro-centrifuge tubes. After addition of 1 µl to 5 µl plasmid solution (e.g. a gateway BP or LR reaction) the mixture was incubated on ice for a few minutes. The entire volume was applied to a electroporation cuvette (0.1 cm gap) which was tapped to ensure the suspension was completely placed between the electrodes and to prevent air bubbles. Using an electroporator a pulse with a field strength of 1.7 kV with a time constant of approximately 4.5 milliseconds was generated and 300 µl (or 10 times the volume of cells) pre-warmed SOC medium (37°C) was immediately added. The entire content of the cuvette was transferred back to a microcentifuge tube and incubated for 1 hour at 37°C at 600-800 rpm shaking to give the cells time to recover. Finally, 50-100 µl of the mixture was applied to selective LB-agar plates and incubated over-night at 37°C.

2.3. Saccharomyces cerevisiae (*S. cerevisiae*)

2.3.1. *S. cerevisiae* strains

L40c: *MATa his3Δ200 trp1-901 leu2-3,112 ade2 lys2-801am can1 URA3*:: (*lexAop*)$_8$-GAL1TATA-*lacZ LYS2*::*(lexAop)$_4$*-HIS3TATA-*HIS3 (Erich Wanker)*

2.3.2. *S. cerevisiae* growth media

1.25x YPD liquid medium

5 g Bacto yeast extract

10 g Bacto peptone

H_20 to a total volume of 400 ml

1.25x YPD agar

5 g Bacto yeast extract

10 g Bacto peptone

10 g Bacto agar

H_20 to a total volume of 400 ml

2.5x Yeast liquid medium (NB)

6,7 g yeast nitrogen base

H_20 to a total volume of 400 ml

1.25x Yeast storage medium (NBG)

3.35 g yeast nitrogen base

250 ml glycerol (99%)

29.44 g betaine

H_20 to a total volume of 400 ml

20x Glucose stock solution

200 g glucose monohydrate

H_20 to a total volume of 500 ml

2.5x agar

10 g Bacto agar

H_20 to a total volume of 200 ml

Amino acid supplements	*Applied concentration*
Leucine	100 mg/ml
Histidine	20 mg/ml
Uracil	20 mg/ml
Tryptophane	25 mg/ml

Amino acid solutions were stored in 100x stock solutions.

2.3.3. Yeast lysis in SDS-loading buffer

Yeast was grown on agar or in liquid culture and processed with a "harsh" lysing method. If grown on agar, entire spots of yeast colonies were stirred directly into SDS-loading buffer using an inoculation loop or if grown in liquid culture, cells were centrifuged for 10 min at 2000 rpm and SDS-loading buffer added (approximately 30 µl buffer per agar spot or pellet resulting from 2 ml liquid culture). Four cycles of 5 min vortexing, 5 min heating to 95°C and subsequent freezing of the samples on dry ice were conducted. Lysate can be readily loaded on a SDS gel for electrophoresis. An alternative method is to incubate cells in NaOH instead of mechanical force to lyse the cell wall.

2.4. Phospho-peptide enrichment from yeast

2.4.1. Reagents

10X Lysis Buffer Mix: 20 mM HEPES, 25 mM sodium pyrophosphate, 10 mM b-glycerophosphate (Cell Signaling Technology Inc., Danvers, MA)

10X IAP Buffer (1x IAP buffer): 50 mM MOPS pH 7.2, 10 mM sodium phosphate, 50 mM NaCl (Cell Signaling Technology Inc., Danvers, MA)

Urea Ultra Pure (Biomol GmbH, Hamburg, Germany)

Orthovanadate (Cell Signaling Technology Inc., Danvers, MA, USA)

DTT (1.25 M) (Cell Signaling Technology Inc., Danvers, MA, USA)

Trypsin-TPCK Solution (1 mg/ml, Cell Signaling Technology Inc., Danvers, MA, USA)

Trypsin MS-grade (Roche Diagnostics GmbH, Mannheim, Germany)

10X HEPES stock solution (200 mM, pH 8) (Cell Signaling Technology Inc., Danvers, MA)

Sep-Pak ® C18 Columns (Waters Corporation Milford, MA, USA)

Phosphotyrosine Mouse mAb (P-Tyr-100) Beads (Cell Signaling Technology Inc., Danvers, MA, USA)

ZipTips C18 Resin (Merck Millipore Corporation, Billerica, MA, USA)

Iodacetamide (Sigma I-6125, St.Louis, MO, USA)

Trifluoracetic acid (AppliChem GmbH, Darmstadt, Germany)

Acetonitrile (Thermo Scientific 51101)

2.4.2. Method

The method was adapted for the use of yeast cells based on the protocol of Rush et al. (2005).

2.4.2.1. Yeast cell lysis and protein digestion

At first, lysis buffer (20 mM HEPES pH 8.0, 9 M urea, 1 mM sodium orthovanadate, 2.5 mM sodium pyrophosphate, 1 mM beta-glycerophosphate) was prepared freshly. Dependent on kinase activity as determined by Western blotting, two to six deep-frozen 2 ml microtubes were placed on ice and zirconia-beads added equal to the volume of the approximately 1ml dry yeast pellet. Subsequently, 500 µl lysis buffer was added and the pellet lysed in a FastPrep24 (MP Biomedicals, Santa Ana, CA) for 20 seconds at highest speed (6.5 Ms^{-1}) homogenizer. The lysate was centrifuged on a cooled table-top centrifuge for 15 min at 15000 rpm at 4°C and the supernatant transferred into a 50 ml centrifugation tube. The lysis was repeated twice adding each time 500 µl lysis buffer. As a next step, 1/10 the volume of obtained supernatant of 45 mM DTT was added, the solution mixed well, and incubated for 20 min at 60°C in a water bath. After cooling the solution to room-temperature (RT) for 10 min, 1/10 the volume of the solution of 110 mM iodoacetamide was applied, mixed well and the incubated for 10 min at RT in the dark. For Trypsin digestion, the solution was diluted 4 times using HEPES buffer to a final concentration of 1 M urea and 10 mM HEPES, pH 8.0. Finally, 1/100 volume of 1 mg/ml trypsin-TPCK solution was added and the proteins in solution digested overnight at RT.

2.4.2.2. C18 column reverse phase chromatography

First, solvent A (0.1% TFA) and solvent B (0.1 % TFA, 40 % acetonitrile (MeCN)) were prepared from 20 % TFA stock solutions. The digest was acidified by addition of 1/20 volume of 20 % TFA solution and incubation for 10 min at RT. In order to remove precipitant the acidified peptide solution was centrifuged for 5 min at 2250 x g and the supernatant decanted into a fresh tube. Subsequently, peptide purification was performed. Therefore, a 10cc reservoir (10 ml syringe reservoir) was connected to the shorter end of the reverse phase chromatography column (Sep-Pak® column). After the column was pre-wetted with 5 ml 100 % MeCN and washed twice with 3.5 ml solvent A, the entire acidified peptide solution was passed though the column by gravity flow. Subsequently, the column was washed by applying 1 ml, then 5 ml, and finally 6 ml of solvent A before eluting the peptides into a polypropylene tube by applying 2 ml solvent B thrice. As a next step, the eluted peptides were freeze-dried two to three days using a CHRIST lyophilizer within the polypropylene tube covered with aluminum foil with poked holes (as incomplete de-gasing may cause the contents to pop out).

2.4.2.3. Immuno-affinity Purification

Lyophilized peptides were spun down and 1.4 ml IAP buffer (Cell Signaling Inc.) was added. To dissolve the peptides the tube was left at RT for 5 min and briefly sonicated in an ultra-sound bath. The pH was controlled to be neutral or not lower than 5 to 6 by spotting 0.5 µl solution on a pH indicator paper (FisherBrand) and the pH adjusted by titrating 1 M Tris Base (not a strong base). All following steps were conducted at 4°C (on-ice or cold-room). The peptide solution was cleared via centrifugation at full-speed for 15 min and transferred directly onto p-Tyr-100 conjugated beads (40 µl, 80 µl slurry) while avoiding bubbles (sample A). The tubes were sealed with laboratory film and incubated on a rotator for 2 hours. After subsequent centrifugation at 2700 x g for 1 minute, the supernatant was transferred to a reaction tube containing 15 µl 4G10 antibody slurry and incubated for 1 hour on a rotator (sample B). The supernatant was removed after centrifugation at 2700 x g for 1 minute. Sample A and B were processed in parallel. Loaded beads were washed two times by addition of 1 ml "IAP buffer plus detergent", inverting the tube 5 times, and removing the supernatant after 1 min centrifugation at 2700 x g. As a next step, the beads were again washed 5 times by applying 1ml purified water, inverting the tube 5 times, and removing the supernatant after 1 min centrifugation at 2700 x g. After the final wash, supernatant was removed until 100-200 µl remained in the tube which was then removed carefully, but completely, by using a smaller volume pipette. Finally, peptides were eluted from the beads by addition of 55 µl of 0.15 % TFA, tapping the bottom of the tube, and letting the tubes stand for 10 minutes at RT. After tapping the bottom of the tube again and brief centrifugation, the supernatant was transferred to a fresh low-peptide-binding tube. Another 45 µl of 0.15% TFA were added to the beads and after tapping the tube and brief centrifugation the supernatants were combined within the low-peptide-binding tube.

2.4.2.4. C18 column peptide concentration and desalting

Both sample A and sample B were processed in parallel. The eluate from the immune-precipitation was centrifuged briefly to collect residual beads on the bottom of the tube. Then, the supernatant was transferred to a fresh tube while ensuring no residual beads were present in the solution which could block the Zip-Tip C18 column. Each eluate was divided into two aliquots of 50 µl each in 1.5 ml low-binding reaction tubes. By attaching a cut 200 µl tip as adaptor to the pipette, the ZipTip was installed while the adaptor was fit tightly into the upper "ring" of the ZipTip, but not further. Setting the pipette to 70 µl, up to 55 µl could be picked up with the tip. C18 columns are processed using solvent A (0.1% TFA) and solvent B (0.1 % TFA, 40 % acetonitrile (MeCN) as previously. For each step, a fresh low-peptide-binding tube was used with the appropriate solvent. First, 40 µl solvent B was aspirated slowly twice into the ZipTip while avoiding air to be pipetted. The columns were

equilibrated twice using two separate tubes by pipetting 50 µl solvent A up and down. The first aliquot was aspirated and expelled carefully 10 times into the tip and the loading repeated with the second aliquot (same tip). Unabsorbed sample was left in the tubes. Subsequently, the tip was washed three times by pipetting up and down 55 µl solvent A. After last washing step the tip was dabbed well on a lint-free tissue. In quick succession, the ZipTip was transferred without adaptor to a 10 µl pipette (set to 10 µl) and peptides eluted in 4 µl solvent B by pipetting three times up and down and one final aspiration into a 0.5 ml non-peptide-binding tube. Finally, the sample was dried in a vacuum concentrator for 60 min.

2.5. Immuno-precipitation of predicted human NRTK targets expressed in yeast

2.5.1. Reagents

Human IgG conjugated agarose beads (Sigma, Taufkirchen, Germany)

Rabbit Anti-goat IgG Horseradish Peroxidase (HRP) Conjugate (Invitrogen GmbH, Darmstadt, Germany)

Lysis buffer	*IP wash solution*
50 mM HEPES-NaOH pH 7.8	50 mM NH_4HCO_3 pH 8
150 mM NaCl	
1 % Triton X	
0.1 % Sodium	
H_20 until total volume reached	

2.5.2. Method

Homology and amino acid sequence motif predicted human kinase targets are verified using an immuno-precipitation protocol in yeast. Human phosphorylation sites were retrieved from phosphositeplus.org database and peptides harboring the site were selected if they were observed in at least three experiments or publications, respectively. Selected human targets were picked from an open reading frame (ORF) collection only if the gateway entry clones do not harbor a stop codon. The "open" entry clones were then cloned via gateway LR reaction into the strong yeast expression vector pRS425_GDP_TAP whereas a Protein A tag from *S. aureus* was C-terminally fused to the predicted human target. After co-transformation of target proteins with the appropriate human kinase (cloned into the pASZ yeast expression vector) into the yeast strain L40c, targets were immuno-precipitated via the Protein A tag on IgG-conjugated agarose beads and tyrosine

phosphorylation events were measured on a Q-Exactive mass spectrometer (Thermo Scientific) and peptides identified using the MaxQuant environment (Cox and Mann, 2008). The method is an adaption of a published chromatin immuno-precipitation protocol of Melanie Grably and David Engelberg (Grably and Engelberg, 2010).

First, the co-transformed yeast is grown in selective medium in two liter cultures which was inoculated with 2% v/v over-night culture. After six hours of growth, expression of human tyrosine kinase was induced by addition of 50 µM $CuSO_4$ and the liter culture was allowed to grow over-night until the highest possible cell density. All subsequent steps were conducted at 4°C to avoid enzymatic activity in the lysate. After centrifugation for 20 min at 2250 x g the yeast pellet was filled into 2 ml centrifugation tubes (with rubber sealed lid) by re-suspending the pellet in residual medium, water or TE buffer (pH 8) in case the solution was to viscous to pipette using a cut tip. The solution was centrifuged again 15 min at 2250 x g and supernatant removed yielding approximately 1 ml dry yeast pellet per centrifugation tube. Thus, five to six milliliter of dry yeast pellet was lysed in a FastPrep24 instrument (settings: 4°C, 6.5 Ms-1, 20 sec) after addition of Zirconia beads equal to the size of the yeast pellet and 500 µl lysis buffer. The lysis was repeated twice adding each time 500 µl lysis buffer and collecting the supernatant in a 15 ml centrifugation tube. Before centrifugation after the third lysis step, the Zirconia beads with 0.5 mm bead radius were removed by poking a 0.4 mm hole using a Bunsen burner heated needle in each 2 ml tube and directly placing it into a 15 ml centrifugation tube using a cut top of a 5 ml syringe reservoir as an adaptor and centrifuged at 2250 x g for 1 min. The bead-free lysate was applied to a sonicator stick and running 5 cycles for 30 sec. The sonicated lysate was centrifuged for 15 min at full speed until the supernatant was clear. Supernatants were combined and residual cell debris removed by centrifuging the approximately 10 ml final lysate at full speed for 15 min and decanting the supernatant. Rabbit IgG-conjugated agarose beads were washed 3 times for 5 min on a rotator by applying 1 ml lysis buffer. 110 µl slurry of IgG beads was added to the combined, cleared 10 ml lysate and incubated over-night in the cold room. Subsequently, the beads were spun down for 1 min at 2000 x g or 2 min at 1500 x g and transferred to a 1.5 ml centrifugation tube. Beads were washed four to six times by addition of 1 ml wash buffer, inverting the tube four times, and removing the supernatant after centrifugation. After carefully removing the supernatant after the last washing step, an equal volume approximately 110 µl (equal volume to bead slurry) of 2.5 x SDS gel loading buffer was added to the beads and the solution heated for 5 min at 95°C. As a next step, the beads and precipitated tagged targets were eluted via the SDS loading buffer and the eluate loaded on a 1 mm thick 10-12 % SDS polyacrylamide gel. The gel ran 15 min at 90 V and 90 min at 110 V (or slower), stained with Coomassie Blue, and visible bands (or the a small area of the gel where the target was expected to reside) excised using a fresh scalpel. Two gel slices were cut out – one slice at the expected molecular weight and one slice above to ensure that slower

running phosphorylated proteins are captured. The gel slices were grinded using a micro-pistil within protein low-binding reaction tubes (LoBind, Eppendorf GmbH) and proteins in-gel digested with trypsin. The resulting peptides were alkylated, reduced and thereafter purified using a C18 column and finally desiccated in a vacuum concentrator.

2.6. Bioinformatics

2.6.1. Methods

2.6.1.1. Amino acid sequence motif generation

The yeast proteome was downloaded 01/06/2011 from SGD (Saccharomyces Genome Database, Stanford, www.yeast-genome.org). The mass spectrometry (MS) output was mapped to the yeast proteome using the SEQUEST algorithm (Eng et al., 1994) and peptides filtered and processed to 15mers where seven amino acids (AAs) are flanking a central tyrosine. For each kinase set a list of 15mers was applied to the iceLogo stand-alone application (Colaert et al., 2009). The background set used was list of 15mers for each tyrosine in the expressed yeast proteome (only of proteins identified in the MS measurements) and "fold change" set as the enrichment/significance parameter. iceLogo outputs motif logo graphics and a comma separated value (CSV) file with the fold change per 15mer position in rows and AA alphabet as columns in a "quasi" position weight matrix (PWM).

2.6.1.2. Motif scoring of phosphorylation sites

Using a custom made PYTHON script all phosphorylation sites were scored according to the logic below. The site score is defined as the sum of all enrichments values over an amino acid 15mer.

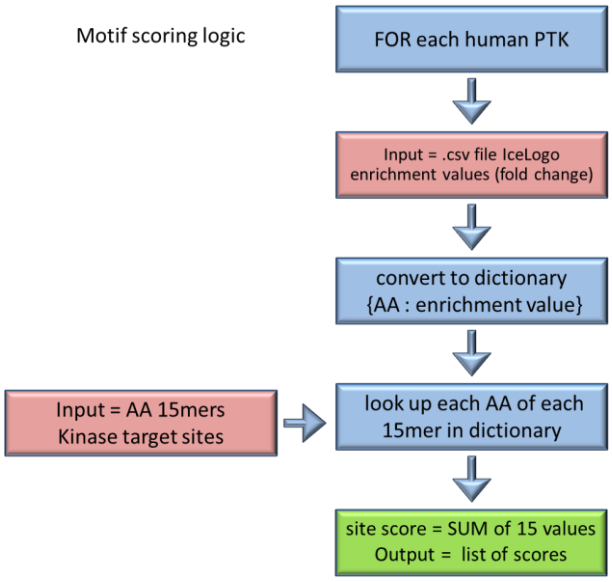


Figure 6: Logical representation of amino acid scoring with input in operations in blue, input in red and output in green.

2.6.1.3. Motif performance testing

R programming language was used for motif performance testing. The module ROCR (Sing et al., 2005) inputs a list of scores with assigned binaries (0 if FALSE, 1 if TRUE) and outputs a graphical display of the performance as ROC (Receiver Operating Characteristic) curve. Thus, the expressed yeast proteome was scored and measured, and targeted sites labeled for each kinase separately. ROCR also outputs accuracy values for each scored site.

2.6.2. Software

Python (www.python.org)

R-project (www.R-project.org)

SnapGeneViewer (SnapGene software from GSL Biotech; www.snapgene.com)

Spyder IDE (https://code.google.com/p/spyderlib/)

2.6.3. Databases

CPDB (Consensus Path Database) (Kamburov et al., 2013)

Inparanoid DB (O'Brien et al., 2005)

PeptideAtlas (www.peptideatlas.org) (Desiere et al., 2006)

PhosphoSitePlus (phosphosite.org) (Hornbeck et al., 2012)

SGD (Saccharomyces Genome Database) (Cherry et al., 2012)

STRING DB (search tool for recurring instances of neighboring genes) (Snel et al., 2000)

2.7. Yeast expression vectors

2.7.1. pASZ vectors

The vectors used for human tyrosine kinase expression in yeast originate from the gateway vector pBTM116 with an implemented copper-inducible yeast promoter from pYEX-BX (*E. coli-S. cerevisiae* shuttle vector; *URA3 LEU2 CUP1* Ap^r ; Clontech Laboratories) and named pASZ. pASZ harbors the gateway cassette with attR recombination sites, cat promoter, Chloramphenicol antibiotic resistance and the control-of-cell-death gene B (ccdB). The gateway cassette is followed by a CYC1 terminator sequence. In one version of the vector (pASZ-CN) a SV40 (Simian Vacuolating Virus 40) nuclear localization sequence (NLS) was cloned between promoter and recombination site leading to a N-

terminally tagged fusion protein. For replication, an origin of replication and lac and T7 promoters are present. Additionally, antibiotic ampicillin- (pASZ-CN) or tetracycline (pASZ-C) resistance and the Phosphoribosylaminoimidazole carboxylase gene (ADE2) with their corresponding promoters as selective markers are encoded in the plasmid. In one version of the vector (pASZ-CN) a SV40 (Simian Vacuolating Virus 40) nuclear localization sequence (NLS) was cloned between promoter and recombination site leading to a N-terminally tagged fusion protein.

2.7.2. pRS425_GDP_TAP

In order to express selected human predicted targets together with human kinases in yeast a strong expressing gateway plasmid was chosen. The pRS425_GDP_TAP harbors a 2µ origin-of-replication for multi-copy transcription flanked selective markers antibiotic ampicillin resistance gene and for selection, the beta-isopropylmalate dehydrogenase (LEU2) gene with promoters. In front of the gateway cassette as described above for pASZ, a strong expressing GAP (glyceraldehyde-3-phosphate dehydrogenase) promoter and also T7 and lac promoters are present. C-terminal adjacent to the gateway cassette a *Staphylococcus aureus* Protein A sequence including a TEV protease cleavage site resulting in Protein A fusion proteins for simple purification using IgG conjugated agarose beads. The Protein A sequence is followed by a C-terminal CYC1 terminator.

2.8. Antibodies

Phospho-tyrosine Mouse Ab “P-Tyr-100” Beads (Cell Signaling Technology Inc., Danvers, MA, USA)

Phospho-tyrosine Mouse Ab “4G10” Beads (Merck Millipore Corporation, Billerica, MA, USA)

Phospho-tyrosine Mouse Ab “4G10” (Merck Millipore Corporation, Billerica, MA, USA)

Goat Anti-Mouse IgG (H+L) Alkaline Phosphatase (AP) Conjugate (Promega Corporation, Fitchburg, WI, USA)

Rabbit Anti-goat IgG Horseradish Peroxidase (HRP) Conjugate (Invitrogen GmbH, Darmstadt, Germany)

ECL Sheep Anti-Mouse Horseradish Peroxidase (HRP) Conjugate (GE Healthcare Europe GmbH, Freiburg, Germany)

Human IgG conjugated agarose beads (Sigma-Aldrich, Taufkirchen, Germany)

2.9. List of chemicals

1-Phenylazo-2-naphthol-6,8-disulfonic acid disodium salt (Orange G) (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Acetic acid (Merck KGaA, Darmstadt, Germany)

Acetone nitrile (Merck KGaA, Darmstadt, Germany)

Acrylamide/Bisacrylamide 40 % (37,5:1) (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Ammonium bicarbonate(Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Ammonium persulfate (Merck KGaA, Darmstadt, Germany)

Ampicillin trihydrate (Sigma-Aldrich Chemie GmbH, Deisenhofen, Germany)

Bacto agar (BD Biosciences, San Jose, CA, USA)

Bacto peptone (BD Biosciences, San Jose, CA, USA)

Bacto tryptone (BD Biosciences, San Jose, CA, USA)

Bacto yeast extract (BD Biosciences, San Jose, CA, USA)

Betain (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Bovine serum albumin fraction V (Roche Diagnostics GmbH, Mannheim, Germany)

Bromphenol blue (Merck KGaA, Darmstadt, Germany)

Calcium chloride dihydrate (Merck KGaA, Darmstadt, Germany)

Coomassie Brilliant Blue G-250 and R-250 (Biomol GmbH, Hamburg)

Dipotassium phosphate (Acros organics part of Thermo Fisher Scientific Inc., Geel, Belgium)

Dithiothreitol (DTT) (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Ethylenediamine tetraacetic acid (EDTA) (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Ethylene glycol tetraacetic acid (EGTA) (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Ethanol (Merck KGaA, Darmstadt, Germany)

Glucose monohydrate (Merck KGaA, Darmstadt, Germany)

Glycerol (Merck KGaA, Darmstadt, Germany)

Glycine (MP Biochemicals, Aurora, USA)

Glycogen (Roche Diagnostics GmbH, Mannheim, Germany)

Histidine (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Isopropanol (Merck KGaA, Darmstadt, Germany)

Kanamycin sulfate (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

LE Agarose (Biozyme Scientific GmbH, Oldendorf, Germany)

Leucine (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Lithiumacetate (LiOAc) (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Magnesium chloride (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Magnesium sulfate (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Methanol (Merck KGaA, Darmstadt, Germany)

Polyethylene glycol (PEG) 3350 (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Polyethylene glycol (PEG) 8000 (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Potassium acetate (Merck KGaA, Darmstadt, Germany)

Potassium chloride (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Sodium carbonate (Merck KGaA, Darmstadt, Germany)

Sodium chloride (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Sodium citrat (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Sodium dihydrogen phosphate (Merck KGaA, Darmstadt, Germany)

Sodium dodecyl sulfate (SDS) (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Sodium hydrogencarbonate (Merck KGaA, Darmstadt, Germany)

Sodium hydroxide (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Sorbitol (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Spectinomycin dihydrochloride pentahydrate (Sigma-Aldrich, Taufkirchen)

Sucrose (Merck KGaA, Darmstadt, Germany)

SYBR Gold Nucleic Acid Gel Stain (Invitrogen GmbH, Darmstadt, Germany)

Tetracycline hydrochloride (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Tetramethylethylenediamine (TEMED) (Invitrogen GmbH, Darmstadt, Germany)

Tris (hydroxymethyl) aminomethane (Tris Base) (Carl Roth GmbH & Co. KG, Karlsruhe, Germany)

Tris (hydroxymethyl) aminomethane hydrochloride (Tris HCl) (Sigma-Aldrich, Taufkirchen)

Triflouroacitic acid

Tryptophan (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Tween 20 (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Uracil (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Yeast nitrogen base (Difco as part of BD Biosciences, San Jose, CA, USA)

2.10. Labware

384well MTPs, PS, flat bottom, clear, sterile, with lid (Greiner Bio-One GmbH, Frickenhausen, Deutschland, 781186)

96well MTPs, PS, flat bottom, crystal clear (Greiner Bio-One GmbH, Frickenhausen, Deutschland, 655101)

96well PCR plate (Costar part of Corning Incorporated, Corning, NY, USA, 6511)

96well deepwell plates (2000 µl/well) (Eppendorf AG, Hamburg, Germany, 0030 501.322)

Agar-plates (241 x 241 x 20) (Nunc GmbH & Co. KG, Wiesbaden, Germany, 240845)

BiomekNX (Beckman Coulter GmbH, Sinsheim, Germany)

Biophotometer plus (Eppendorf AG, Hamburg, Germany)

Biorad DC Protein Assay (Bio-Rad Laboratories, Hercules, CA, USA)

Centrifuge 5810 R (Eppendorf AG, Hamburg, Germany)

Chromatography Paper 3mm (Whatman International Ltd., Maidstone, England, UK)

E.A.S.Y 429k digital camera (Herolab GmbH Laborgeräte, Wiesloch, Germany)

Glass beads, acid-washed <106 µm Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany, G4649)

Incubator 1000 (Heidolph Instruments GmbH & Co. KG, Schwabach, Germany)

Intas Gel-documentation G (Intas Science Imaging Instruments GmbH, Göttingen, Germany)

Lyophilizer CHRIST ALPHA1-2, Martin Christ Gefriertrocknungsanlagen GmbH, Osterode am Harz, Germany)

Mini-protean tetra cell electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA)

Micro-pistils (VWR International, Radnor, PA, USA)

NanoDrop 2000 (Thermo Fisher Scientific Inc., Waltham, MA, USA)

Nitrocellulose membrane (Bio-Rad Laboratories, Hercules, CA, USA, 162-0115)

Omnitrays (Nunc GmbH & Co. KG, Wiesbaden, Germany, 165218)

Pin tools with 96 and 384 pins. The steel pins are cylindrical with a diameter of 1.3 mm and the edge of the flat top that is touching the agar is beveled 45°at 0.2 mm.

Polyvinylidene fluoride (PVDF) membrane (Bio-Rad Laboratories, Hercules, CA, USA, 162-0177)

PowerPac universal power supply (Bio-Rad Laboratories, Hercules, CA, USA)

Protein LoBind reaction tubes (Eppendorf AG, Hamburg, Germany)

Sealing tape for PCR plates/MTPs (Costar part of Corning Incorporated, Corning, NY, USA, 6524 or Thermo Fisher Scientific Inc., Waltham, MA, USA, AB-5558)

Sterile breathable sealing films (Aeraseal, Excel Scientific Inc., Victorville, CA, USA, BS-25)

Sunrise 96 horizontal gel electrophoresis apparatus (Biometra GmbH, Göttingen, Germany)

Tetrad PTC-225 thermo cycler (MJ Research Inc., St. Bruno, Canada)

Thermomixer comfort (Eppendorf AG, Hamburg, Germany)

Titramax 1000 (Heidolph Instruments GmbH & Co. KG, Schwabach, Germany)

Trans-blot SD semi-dry transfer cell (Bio-Rad Laboratories, Hercules, CA, USA)

Zirconia beads (Carl Roth GmbH & Co. KG, Karlsruhe, Germany, N034.1)

2.11. Enzymes and reagents

1 Kb Plus DNA ladder (Invitrogen , Carlsbad, CA, USA)

AttoPhos Substrate Set (Roche Diagnostics GmbH, Mannheim, Germany)

BP Clonase Enzyme Mix (Invitrogen , Carlsbad, CA, USA)

dNTP-Mix (Fermentas GmbH, St. Leon-Rot, Germany)

FastDigest Bsp1407I (Fermentas GmbH, St. Leon-Rot, Germany)

KOD polymerase (Novagen/Toyoba, Merck Millipore Corporation, Billerica, MA, USA)

LR Clonase Enzyme Mix II (Invitrogen , Carlsbad, CA, USA)

PfuTurbo DNA Polymerase (Stratagene, Santa Clara, USA)

Phusion Hot Start High-Fidelity DNA Polymerase (Finnzymes/ Thermo Fisher Scientific, Vantaa, Finland)

Prestained protein ladder, PageRuler™ Plus (Fermentas GmbH, St. Leon-Rot, Germany)

Proteinase K Solution (Invitrogen , Carlsbad, CA, USA)

QIAprep Spin Miniprep kit (Qiagen GmbH, Hilden, Germany)

QIAquick PCR Purification Kit (Qiagen GmbH, Hilden, Germany)

Salmon sperm Carrier DNA (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany)

Trypsin MS grade (Roche Diagnostics GmbH, Mannheim, Germany)

Trypsin-EDTA (Gibco BRL, Gaithersburg, MD, USA)

Western Lightning Plus-ECL (PerkinElmer Inc., Waltham, MA, USA)

3. Results

3.1. Discovery of human NRTK activity in yeast

Initially, human tyrosine kinases were expressed in yeast in order to screen for tyrosine phosphorylation dependent human protein-protein interactions using yeast-2-hybrid technology (Grossmann et al., 2015). All human cytoplasmic PTKs coding sequences were cloned into pASZ, a copper inducible yeast expression vector, with and without an N-terminal nuclear localization sequence (NLS) and expressed in yeast strains (Material and methods). Minor fitness defects of the yeast due to controlled, low NRTK overexpression indicated the strains could be useful tools to systematically investigate specificity and activity of human NRTKs. Western blot analysis using the highly sensitive pan anti-phospho-tyrosine antibody 4G10 showed variable phosphorylation of yeast proteins depending on the kinase (Figure 8). For 26 out of 59 clones tested, comprising 16 out of 32 of all currently annotated cytoplasmic tyrosine kinases (Blume-Jensen and Hunter, 2001), the activity signal was clearly distinct from background and hence regarded as sufficiently strong for further analysis by the approach presented here. Tyrosine kinases were previously assigned to kinase families according to their domain structure such as the placement of regulatory domains including proline-rich motif (PxxP) recognizing SRC-homology 3 (SH3) domain and phospho-tyrosine binding SRC-homology 2 (SH2) domain with respect to the conserved kinase domain (Blume-Jensen and Hunter, 2001). As it can be seen in Figure 7**,** at least one member of each kinase family was found to be active in yeast providing a representative set for this group of enzymes for analysis. Interestingly, all SRC-family kinases (SFKs), with the exception of Lymphocyte cell-specific protein-tyrosine kinase LCK, showed phosphorylation activity suggesting similar regulatory mechanisms within this kinase family and/or sufficient expression in yeast (Boggon and Eck, 2004, Bradshaw, 2010). Oppositely, BMX was the only TEC family kinase showing activity in yeast which may be due to its unique structure within this family of NRTKs (Bradshaw, 2010).

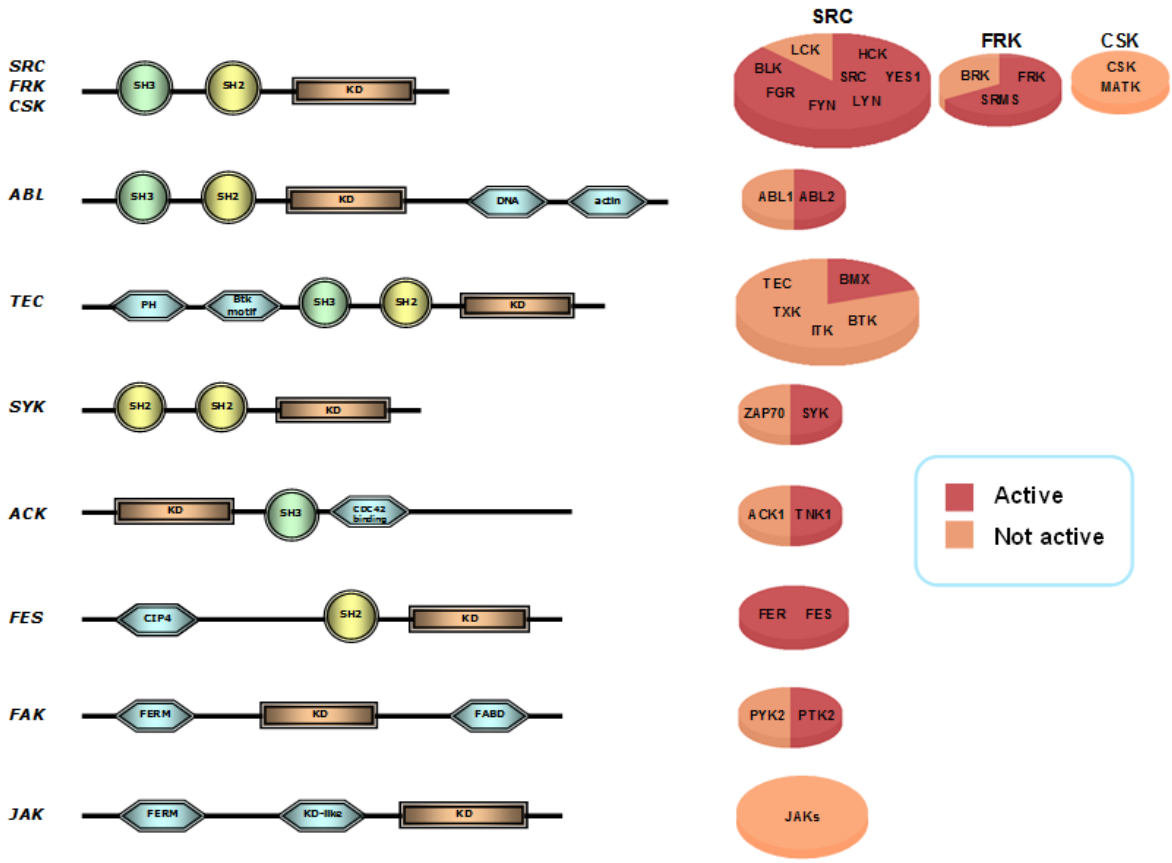


Figure 7: Human cytoplasmic protein tyrosine kinase (NRTK) families assigned according to their domain architecture (modified from Blume-Jensen and Hunter (2001)) and pie charts indicating coverage of active NRTKs in yeast.

Overall 24 active full-length human NRTKs were expressed in individual yeast strains whereas focal adhesion kinase (FAK or PTK2) and FYN (Table 2) were modified. For these two kinases N-terminal inhibitory domains, i.e. the FERM domain and N-and C-terminal regulatory sequences (SH2 and SH3 domains, and an inhibitory C-terminal loop), were truncated. This rendered FAK active in yeast and hyper-activated FYN. The truncated version of FYN is regarded as a separate kinase in this work. In human cells, PTKs are under tight regulation bearing inhibitory and activating phosphorylation sites and regulatory domains (Blume-Jensen and Hunter, 2001). On the contrary, *S. cerevisiae* does not contain bona-fide tyrosine kinases and any trans-regulatory components are not in place. Western blotting permits a comparison in phosphorylation activity between kinase samples by anti-phospho-tyrosine staining (Figure 8).

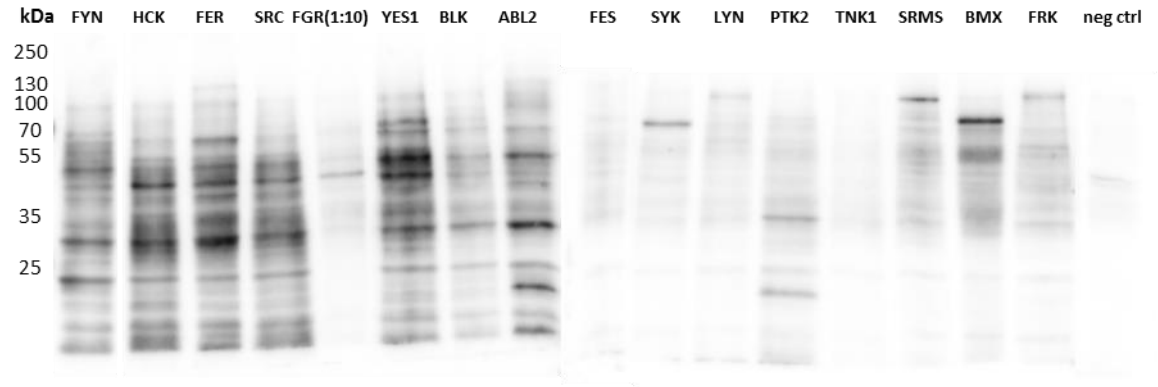


Figure 8: Western blot analysis using 4G10 anti-phospho-tyrosine antibody of yeast lysates sampled from the material used for the enrichment procedure. Samples were mixed 1:1 for clones with and without NLS when both clones showed activity. All samples were equally loaded as controlled for by visual inspection using Coomassie Blue staining (picture not available). Truncated FYN was not loaded and FGR (showing the second strongest Western blot signal) and was loaded 1:10 as the camera adjusts for highest signal intensity and hence weaker signals would be masked.

Activity, and to a certain extent phosphorylation specificity, can be observed in the appearance of kinase preparation specific bands with different relative intensities. Strongest antibody signals were detected for the SRC-kinases, ABL2, and FER in the following order(strongest to weakest): Truncated FYN construct, FGR, YES1, HCK, FER, ABL2, FYN, SRC, BLK. The kinases showing weaker anti-body staining were BMX, FRK, SYK, LYN, PTK2, SRMS, FES, TNK1 (Table 3)**.**

The expression of active enzymes in a model organism should not perturb normal growth and homeostasis in order to mimic the normal cellular environment. Therefore, it was necessary for each individual kinase to find the adequate copper concentration as inducer for expression. Thus, it was ensured that neither the kinase nor the copper itself causes growth phenotypes or harm the yeast in other ways, respectively, while obtaining the highest possible phosphorylation activity. In order to control for low toxicity assay conditions all yeast preparations were plated in a series of 10-fold dilutions on selective media with varying copper concentrations ranging from 50 µM to 6.4 mM. Copper concentrations above 100 µM diminished yeast growth when expressing NRTKs showing strong phosphorylation activity such as ABL2 and HCK. This suggested that those kinases themselves are toxic to yeast cells at this level of induction due to the fact that lower active kinase transformed yeast grew normally under the same condition. In agreement with reports of tyrosine kinase toxicity in yeast (Superti-Furga et al., 1993, Sekigawa et al., 2010) all yeast colonies appeared smaller at copper concentrations 400 µM and 800 µM whereas no growth was detected at the next higher concentration tested (1.6 mM). Hence it can be hypothesized the copper itself is harming the yeast at 400 µM and preventing growth of the organism at a concentration above 800 µM. Chosen copper concentrations for each clone are denoted with a star in Figure 9 where all 26 yeast transformations are plated in 1:10 dilutions on 20, 50, and 100 µM copper sulfate containing selective agar. None of

the yeast colonies showed apparent growth phenotypes at concentrations used for further analysis suggesting low or absent toxicity caused by the kinase or the heavy metal.

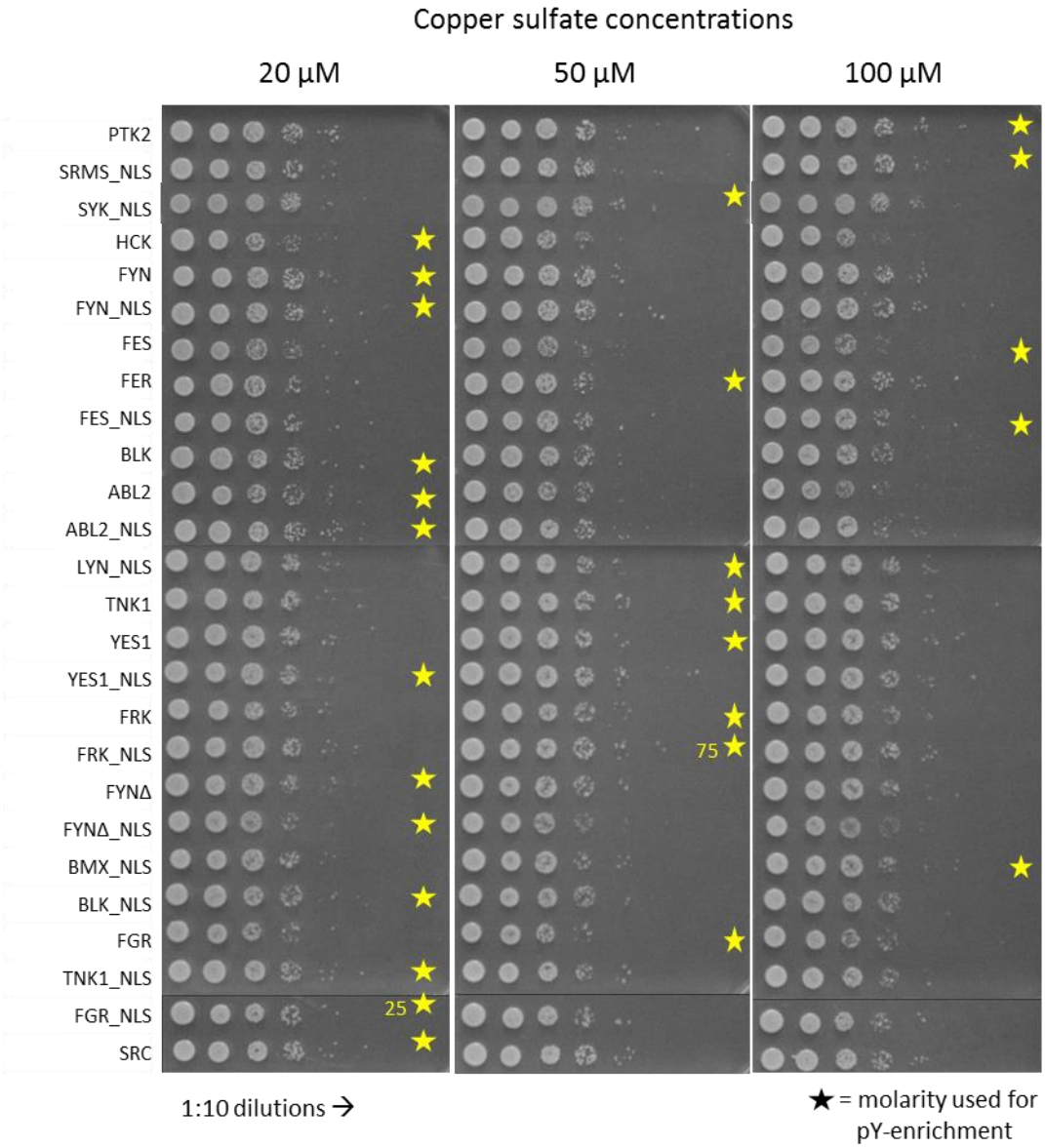


Figure 9: Toxicity test. Human NRTKs were expressed and yeast was spotted in 1:10 dilutions on selective medium with copper concentrations as indicated and grown for 2 days.

Gene symbol	Entrez Gene ID	RefSeq ID	CloneID (ProtID)	Clone description	Kinase full name
ABL2	27	NM_005158	IOH40442	cds AA 1-1167 full-length	Abelson murine leukemia viral oncogene homolog 2 or Abelson-related gene (Arg)
BLK	640	NM_001715	CCSB_5424	cds AA 1-505 full-length	B lymphocyte kinase
BMX	660	NM_203281	IOH11645	cds AA 1-675 full-length	Bone marrow tyrosine kinase gene in chromosome X protein
FER	2241	NM_005246	OCAB_100014538	cds AA 1-822 full-length no stop: plus "SGLMGPAFLYKVVSNstop"	Fujinami poultry sarcoma/Feline sarcoma-related protein Fer
FES	2242	NM_002005	RZPDo839H03100	cds AA 1-822 full-length	Feline sarcoma/Fujinami avian sarcoma oncogene homolog
FGR	2268	NM_005248	FLJ26678AAAN	cds AA 1-528 full-length	Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog
FRK	2444	NM_002031	RZPDo839E01123	cds AA 1-505 full-length	Fyn-related kinase or Nuclear tyrosine protein kinase RAK
FYN	**2534**	NM_002037	IOH21081	cds AA 1-537 full-length	FYN oncogene related to SRC FGR YES or src/yes-related novel gene (SYN)
FynTr	2534	NM_002037	FYN	AA 258-520 ; N-terminal SH3+SH2 and C-terminal 17AA removed	FYN oncogene related to SRC FGR YES or src/yes-related novel gene (SYN)
HCK	3055	NM_002110	CCSB_9159	cds AA 1-526 full-length	Hematopoietic/Hemopoietic cell kinase
LYN	4067	NM_002350	CCSB_54561	cds AA 1-512 full-length	Lck/Yes-related novel protein tyrosine kinase
PTK2	5747	NM_001199649	CCSB_1682	AA 1-392 (FERM domains) and 868-878 and 915-917 and 919 missing	Protein-tyrosine kinase 2 or Focal Adhesion Kinase 1 (FAK)
SRC	6714	NM_005417	CCSB_7119	cds AA 1-536 full-length	("Sarcoma") proto-oncogene c-Src
SRMS	6725	NM_080823	OCAB_100015619	cds AA 1-488 full-length no stop: plus "SGLMGPAFLYKVVSNstop"	src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristoylation sites
SYK	6850	NM_003177	CCSB_3739	cds AA 1-635 full-length	Spleen tyrosine kinase
TNK1	8711	NM_001251902	RZPDo839E0798	AA 411-415 "SSSFH" missing, but full-length otherwise cds AA 1-666	tyrosine kinase non-receptor 1 or CD38 negative kinase 1
YES1	7525	NM_005433	CCSB_10705	cds AA 1-543 full-length	Yamaguchi sarcoma oncogene homolog 1

Table 2: Clone descriptions

Gene symbol	Activity (WB signal)	No. sites identified	MS spectral counts
ABL2	++++	176	603
ABL2_NLS	++++	144	528
BLK	++	96	202
BLK_NLS	++++	218	709
BMX_NLS	+++	119	444
FER	++++	266	958
FES	+	16	39
FES_NLS	++	78	213
FGR	+++++	326	1285
FGR_NLS	++++	239	734
FRK	+++	226	672
FRK_NLS	++	62	135
FYN	++++	144	394
FYN_NLS	++++	119	333
FYNΔ_NLS	+++++	274	1111
FYNΔ	+++++	327	1507
HCK	++++	219	659
LYN_NLS	+++	155	556
PTK2	++	124	419
SRC	++++	108	479
SRMS_NLS	++	28	55
SYK_NLS	+++	96	274
TNK1	++	206	825
TNK1_NLS	+	254	872
YES1	+++	210	800
YES1_NLS	++++	87	345

Table 3: Activity of employed tyrosine kinase clones in terms of Western blot signal ("+"=weak; "+++++"=very strong), number of identified sites and sum of yeast target peptide fragment spectral counts.

3.2. Human NRTK expression and sample preparation

Yeast expressing human tyrosine kinases were grown in liter cultures and processed applying an adapted phospho-tyrosine enrichment protocol from Cell Signaling Technology Inc. (Figure 10). A major advantage of using yeast is the virtually unlimited amount of material for analysis from liquid cultures. Hence, two to six vials containing 1 ml dry yeast pellet were lysed in a buffer containing 9 M Urea. After diluting the urea containing cell lysate denatured proteins were digested with Trypsin over-night to cleave C-terminal to arginine and lysine residues. Peptides were purified using reverse phase chromatography C18 columns. Subsequently, peptide enriched solutions were freeze-dried via

lyophilization and tyrosine-phosphorylated peptides enriched in a two-step immuno-precipitation (IP) using two different anti-phospho-tyrosine antibodies conjugated sepharose beads consecutively. Most of the modified peptides were captured in the first IP using pTyr-100 antibody-conjugated beads. A subsequent complementary second IP using 4G10 antibody-conjugated beads was performed on the flow-through to capture modified peptides not bound in the first IP (less than 10% of all measured sites). Finally, the off-bead elution was desalted and modified peptides concentrated on an additional C18 column ("Zip-tips") and liquids removed using a vacuum-centrifuge. Having the advantage of unlimited yeast material and successful adaption of a commercially available enrichment protocol resulted in a high number of measured enriched pY-peptides compared to samples from human cell culture.

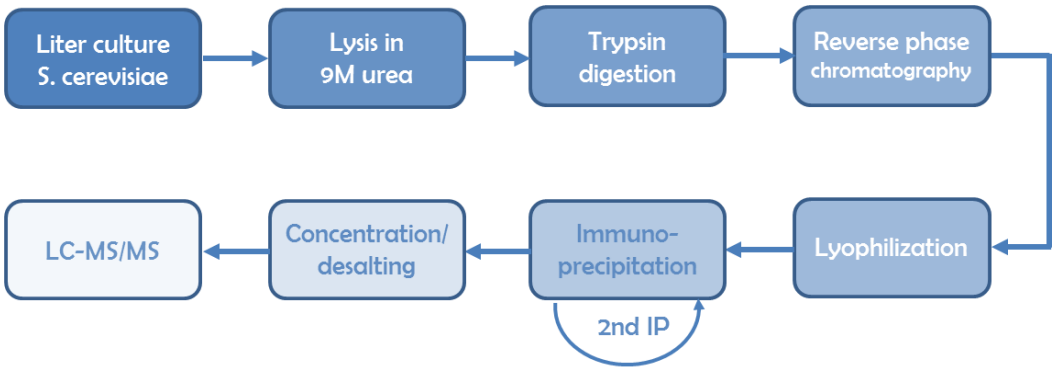


Figure 10: Experimental workflow for phospho-tyrosine peptide enrichment (adapted from Cell Signaling Technology Inc. (Rush et al., 2005))

3.3. Determination of tyrosine phosphorylation sites by mass spectrometry

Mass spectrometry (MS) measurements were performed on a total of 60 phospho-tyrosine peptide samples from 26 different yeast preparation as described in detail the Material and Methods section. Dry peptide enrichments were shipped and measured by Prof. Dr. Bryan Ballif, University of Vermont, USA, as previously (Cheerathodi et al., 2015) using a MicroAs autosampler, a Surveyor PumpPlus HPLC, and a linear ion trap-orbitrap (LTQ-Orbitrap) platform (Thermo Electron, Waltham, MA, USA). Peptides were mapped to the yeast proteome (downloaded 01/06/2011 from SGD) using the SEQUEST algorithm (Eng et al., 1994). Initially, the mapping software searches an amino acid sequence database for sequences matching the mass of the measured peptides. Subsequently, the matched amino acid sequences are converted to fragmentation pattern and compared to the MS/MS spectra yielding a preliminary score. Each of the measured spectra of the best 500 preliminary scoring peptides is correlated to theoretical spectra (predicted mass over charge values for fragment

ions originating from *in silico* peptide library fragmentation). Thus, a cross-correlation value ("XCorr") is retrieved from comparing fragment ion patterns which enables ranking of best matched candidate peptides and circumvents the need for manual MS/MS spectra inspection. Usually, a XCorr value of 2.0 indicates a good correlation and was chosen as cut-off. Additionally, a "delta correlation" value ("Delta Corr") indicates how different the first hit is from the second hit in the search results whereas a value of above 0.1 is considered good (as a rule of thumb). Additionally, a mass accuracy value is provided showing the difference in expected to observed peptide mass in parts per million (ppm). The algorithm Ascore (Beausoleil et al., 2006) was used in order to determine the presence or absence of modification site-spanning fragmentation ions. The program outputs a homonymous score "Acore" which is a probability measure for correct phosphorylation site localization based on the presence and intensity of phosphorylation site-determining ions in MS/MS spectra. Thus, the raw output from SEQUEST algorithm provides the scan number, sample name, peptide charge, peptide sequence, matched protein identifier, mass accuracy value, XCorr, DeltaCorr, redundancy, and Ascore for each modification site. The output was initially filtered according to technical (mapping) performance according to the parameter described above and the resulting list of candidates further filtered according to biological criteria and requirements as described in the next section.

3.4. Mass spectrometry data filtering and normalization

The received peptide data from mass spectrometry measurements was initially filtered by removing false positives (matches against the decoy sequences), S/T phosphorylated peptides, and hits mapped to yeast background tyrosine phosphorylation as reported in SwissProt database (UniProt, 2015) and two yeast proteome screens (Gnad et al., 2009, Tan et al., 2009a). For 22 peptides with two adjacent tyrosine residues on a single peptide the N-terminal site was chosen as the mapping software could not define the distinct modification site localization due to the same mass of both possible peptide instances.

A total of 900 different proteins were measured to be targeted in yeast. One third of these proteins and approximately half of 1433 measured modification sites in total was targeted by a only single kinase (Figure 11) indicating human kinase specificity was preserved in the yeast preparations. The distributions of kinases per protein and kinases per site tail-out to a maximum of 16 and 17 kinases per protein and site, respectively. Interestingly, tyrosine residues targeted by nine or more kinases showed strongly increasing average intensity values with increasing number of kinase per site (Figure 11). This might be due to a bias in mass spectrometry measurements when particular peptides have optimum ionization and fragmentation properties or due to unknown auto-phosphorylation events in yeast.

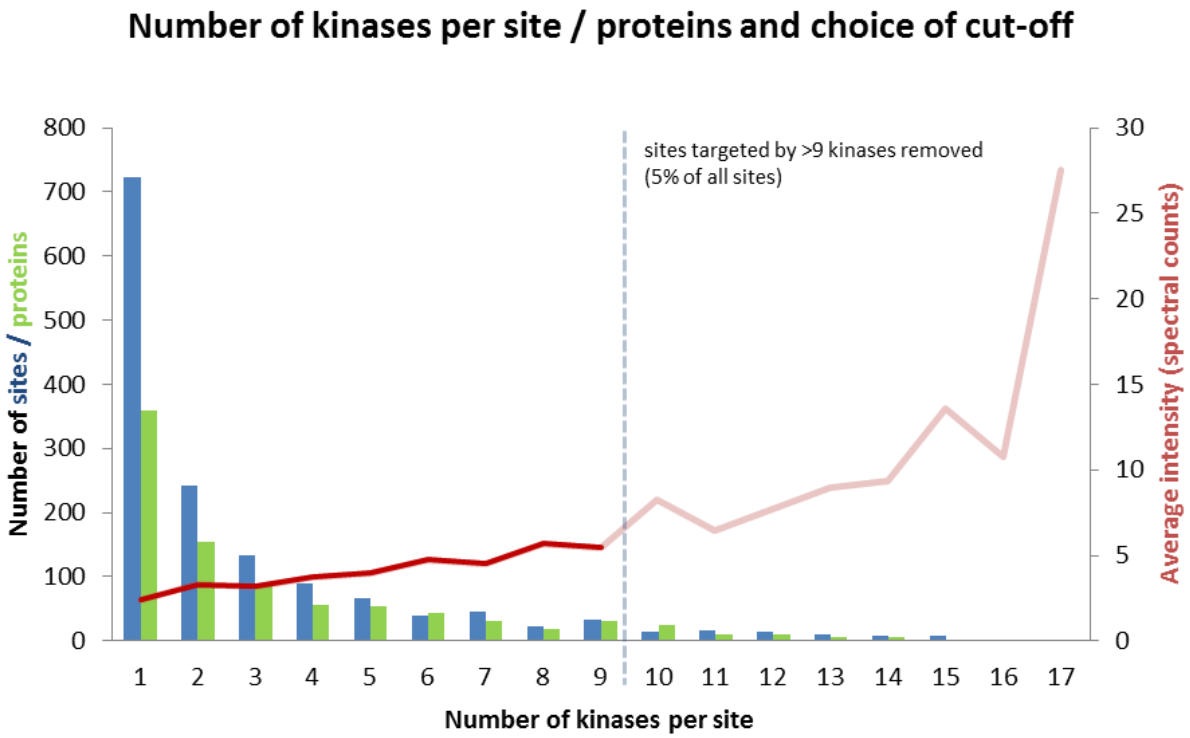


Figure 11: Bar graph showing the number of yeast phosphorylation sites and corresponding number of yeast proteins (left vertical axis) and the average intensity (average spectral counts, right vertical axis) per phosphorylation site plotted against the number of targeting human NRTKs. The dashed line indicates the chosen cut-off excluding sites targeted by 10 sites or more per kinase (approximately 5% of all targeted sites).

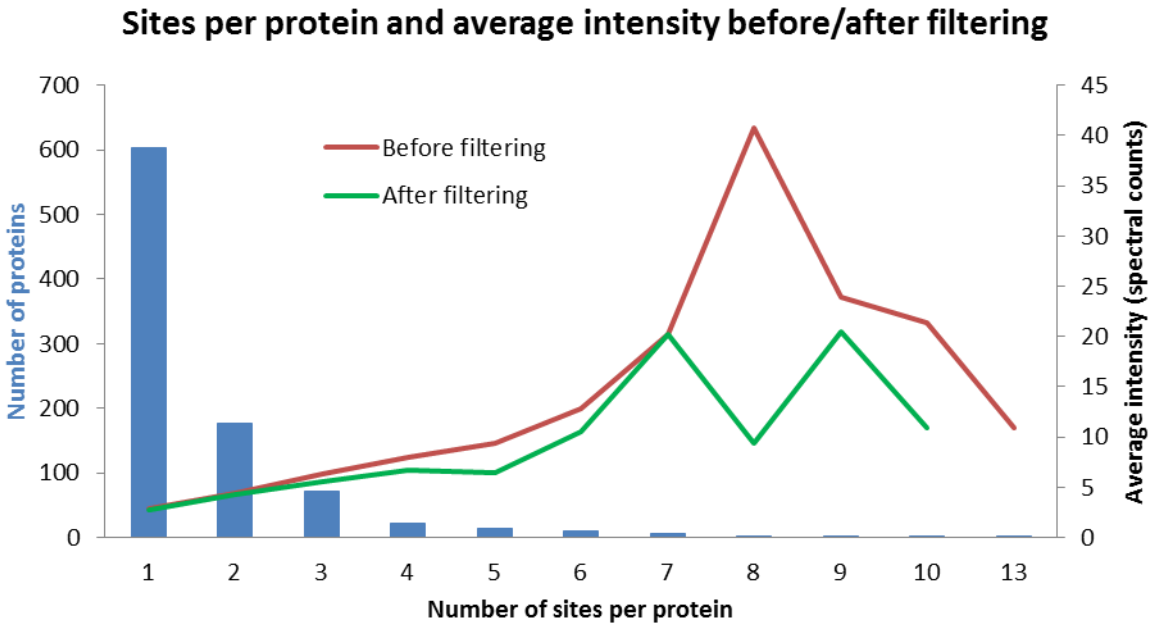


Figure 12: Bar graph representing the number proteins modified on 1 to 13 tyrosine residues. The line graph shows the average intensity distribution before (red line) and after (green line) the chosen cut-off at nine kinases per site.

Moreover, residues on proteins targeted by many or all kinases were highly abundant in yeast being part of the Ribosome, Proteasome, and other conserved protein complexes. This trend can also be observed when plotting the number of pY-sites per protein as shown in Figure 12 in which the average intensity increased with the number of sites per protein until a peak at 8 sites per protein. Hence, we chose to filter sites targeted by more than nine of the 16 kinases in total comprising five percent of all sites identified. As a result, average intensity measures for the number of sites identified per protein showed a more equal distribution after filtering (Figure 12) supporting the choice of cut-off. The peak of average intensity values at 8 sites per protein disappeared leaving a maximum of 22 average spectral counts per proteins for the proteins targeted on seven sites or more which however, are only few. After filtering, the number of sites per kinases varied from 29 sites for SRMS to 464 sites for FGR showing least and most phosphorylation activity in yeast, respectively (Figure 13). The average ratio of sites over proteins targeted was 1.21 among all kinases. Some kinases such as FGR, PTK2, and TNK1 appeared to have targeted more sites on fewer proteins in comparison to other kinases such as FES, LYN, and SRC (Figure 13) whereas overall the ratios were rather equal for all NRTKs.

After initial filtering we analyzed whether protein abundance played a significant role in detection, i.e. whether rather highly expressed yeast proteins were targeted by the human kinase or were preferentially measured by the mass spectrometry instrument. For this purpose, three datasets were retrieved reporting whole proteome quantifications in yeast. In 2003, Huh et al. created a green fluorescent protein (GFP)-tagged yeast mRNA library and was hence able to monitor localization and mRNA expression of approximately 75% of the yeast proteome. In 2006, Newman et al. used high-throughput flow cytometry and a library of GFP-tagged yeast strains to detect protein levels at single-cell resolution. Moreover, in 2008 de Godoy et al. performed a comparison of haploid versus diploid yeast proteomes by SILAC quantification using a mass spectrometry read out.

The analysis of protein abundance was primarily conducted by the former Master student Federico Apelt (Apelt, 2012) and hence the outcomes are presented only in brief. Approximately 44 percent of the detectable yeast proteome was measured in all 3 reference studies while the method of Godoy et al. captured rather high abundant, Huh et al. medium range abundant, and Newman et al. rather low abundant yeast proteins. Hence, abundances were standardized by combining the three sets. The intensity distributions of proteins measured in the MS experiments presented here were similar to the distributions of the corresponding proteins in each of the other three studies and the combined abundance dataset. Importantly, plotting the standardized protein abundance against the intensity values per kinase suggests that there is no dependence of the number of the measured spectral counts on relative protein abundance in yeast (Apelt, 2012). In order to standardize the data

to kinase activity we tested the normalization of intensity values towards maximum intensity per kinase, summed intensities per kinase, and total number of phosphorylation sites (pY-sites) per kinase. The log frequency distribution was plotted against the log intensity divided by one of the three parameter above. Applying the number of pY-sites per kinase led to most similar intensity profiles between kinases and hence this parameter was chosen for normalization (Apelt, 2012).

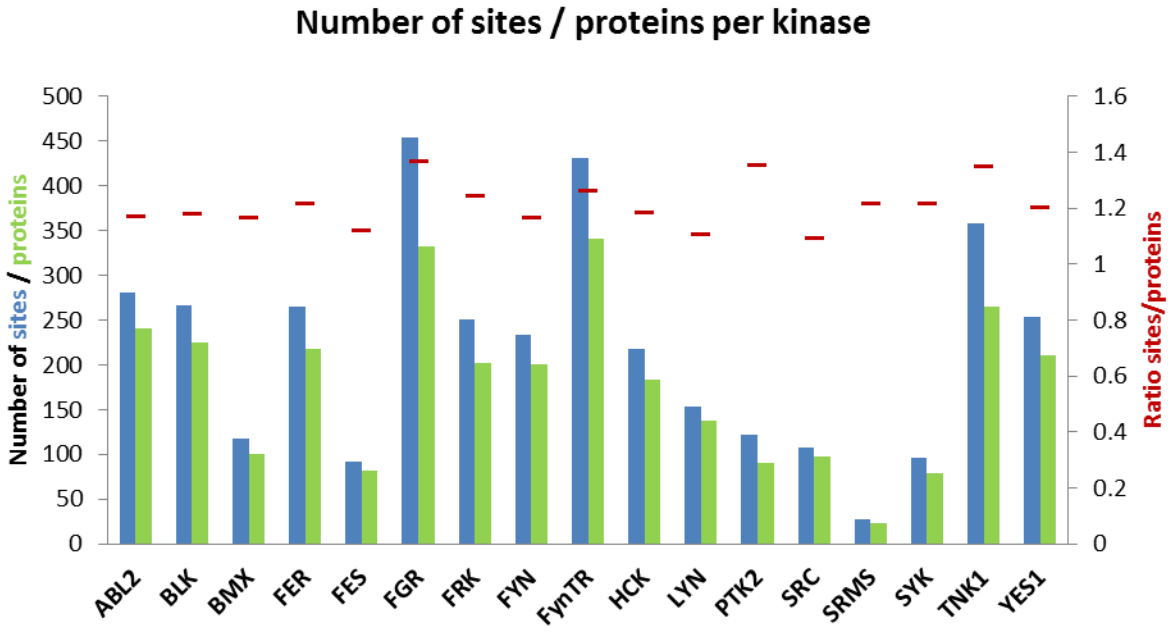
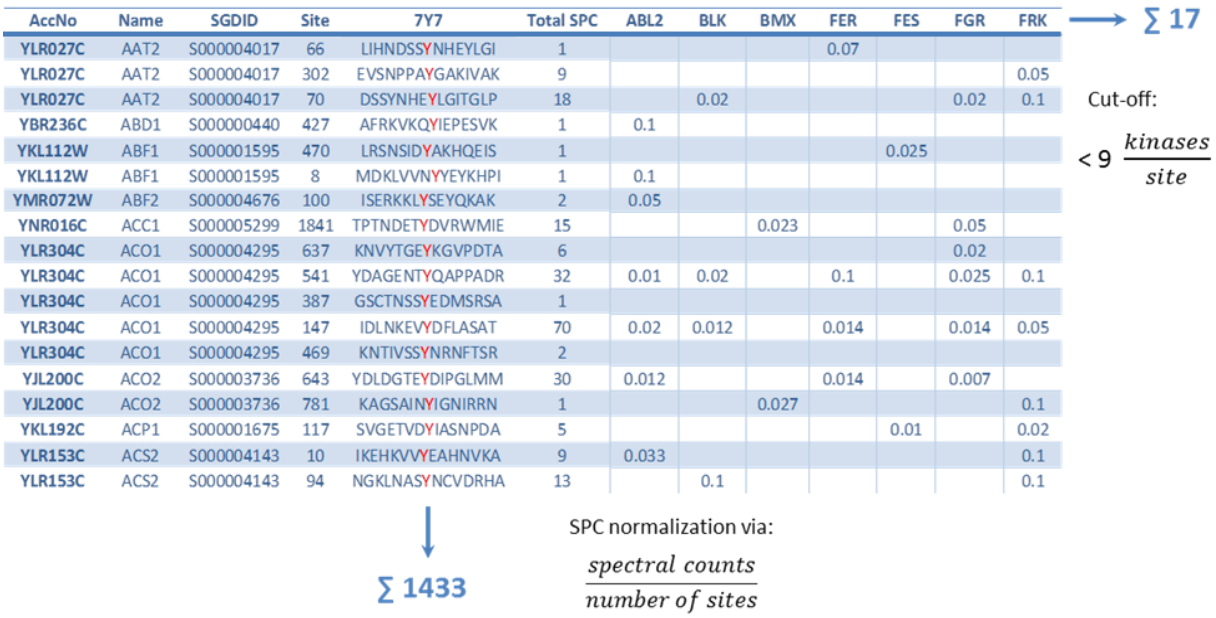


Figure 13: Bar graph showing the number of yeast phosphorylation sites and targeted yeast proteins per human NRTK (preparations for NRTKs with and without NLS combined). The left vertical axis shows the number of target sites and proteins and the right vertical axis the ratio of sites over proteins.

After filtering and normalization an overall number of 15276 tyrosine phosphorylated peptides from 60 MS samples were retrieved representing 1433 unique tyrosine phosphorylation events on 900 yeast proteins. Figure 14 depicts a representation of the retrieved data matrix with six columns harboring identifiers, sequence information, and total spectral counts for each phosphorylation site in 1433 rows and further 17 columns: one for each kinase preparation with normalized spectral counts as values for each phosphorylation site if targeted. The spectral counts as matrix values were normalized by the number of kinases per site and by division of the total number of sites per kinase (as a measure of kinase activity) as described above.



AccNo	Name	SGDID	Site	7Y7	Total SPC	ABL2	BLK	BMX	FER	FES	FGR	FRK
YLR027C	AAT2	S000004017	66	LIHNDSSYNHEYLGI	1				0.07			
YLR027C	AAT2	S000004017	302	EVSNPPAYGAKIVAK	9							0.05
YLR027C	AAT2	S000004017	70	DSSYNHEYLGITGLP	18		0.02				0.02	0.1
YBR236C	ABD1	S000000440	427	AFRKVKQYIEPESVK	1	0.1						
YKL112W	ABF1	S000001595	470	LRSNSIDYAKHQEIS	1					0.025		
YKL112W	ABF1	S000001595	8	MDKLVVNYYEYKHPI	1	0.1						
YMR072W	ABF2	S000004676	100	ISERKKLYSEYQKAK	2	0.05						
YNR016C	ACC1	S000005299	1841	TPTNDETYDVRWMIE	15			0.023			0.05	
YLR304C	ACO1	S000004295	637	KNVYTGEYKGVPDTA	6						0.02	
YLR304C	ACO1	S000004295	541	YDAGENTYQAPPADR	32	0.01	0.02		0.1		0.025	0.1
YLR304C	ACO1	S000004295	387	GSCTNSSYEDMSRSA	1							
YLR304C	ACO1	S000004295	147	IDLNKEVYDFLASAT	70	0.02	0.012		0.014		0.014	0.05
YLR304C	ACO1	S000004295	469	KNTIVSSYNRNFTSR	2							
YJL200C	ACO2	S000003736	643	YDLDGTEYDIPGLMM	30	0.012			0.014		0.007	
YJL200C	ACO2	S000003736	781	KAGSAINYIGNIRRN	1			0.027				0.1
YKL192C	ACP1	S000001675	117	SVGETVDYIASNPDA	5					0.01		0.02
YLR153C	ACS2	S000004143	10	IKEHKVVYEAHNVKA	9	0.033						0.1
YLR153C	ACS2	S000004143	94	NGKLNASYNCVDRHA	13		0.1					0.1

Figure 14: Data representation as a matrix having six columns harboring identifiers, sequence information and total spectral counts for each phosphorylation site in 1433 rows and 17 columns, one for each kinase preparation with normalized spectral counts as values for each phosphorylation site if targeted.

3.5. Comparative analysis of yeast substrate targeting

In order to pairwise compare the kinase in substrate targeting the interception over union of shared targets is visualized in Figure 15. By plotting pairwise Jaccard indices using the normalized intensities and clustering according to similarity we could show known and novel relationships. Hierarchical clustering was performed by row averages leading to a matrix symmetric at the diagonal. As expected, the SRC-family kinases (SFKs) YES1, SRC, FYN, FGR, and HCK cluster together (red box). Furthermore, the SFKs LYN and BLK share many pY-sites with the other SFKs, however, form a distinct cluster (violet box) with ABL2 and BMX. BMX and BLK both have a strong target overlap with LYN, but less between each other. Unexpectedly, FER which together with FES resembles a separate kinase family, shares many targets with the FRK family (orange box) and to some extent with SFKs. SRMS and SYK have little overlap to all other kinases. It is noteworthy that SRMS was the least active human PTK in yeast and comparisons are likely less significant because of the small number of 28 non-redundant sites measured. Interestingly, PTK2 and TNK1 which are members of different kinase families with differing domain architecture (green box) cluster together.

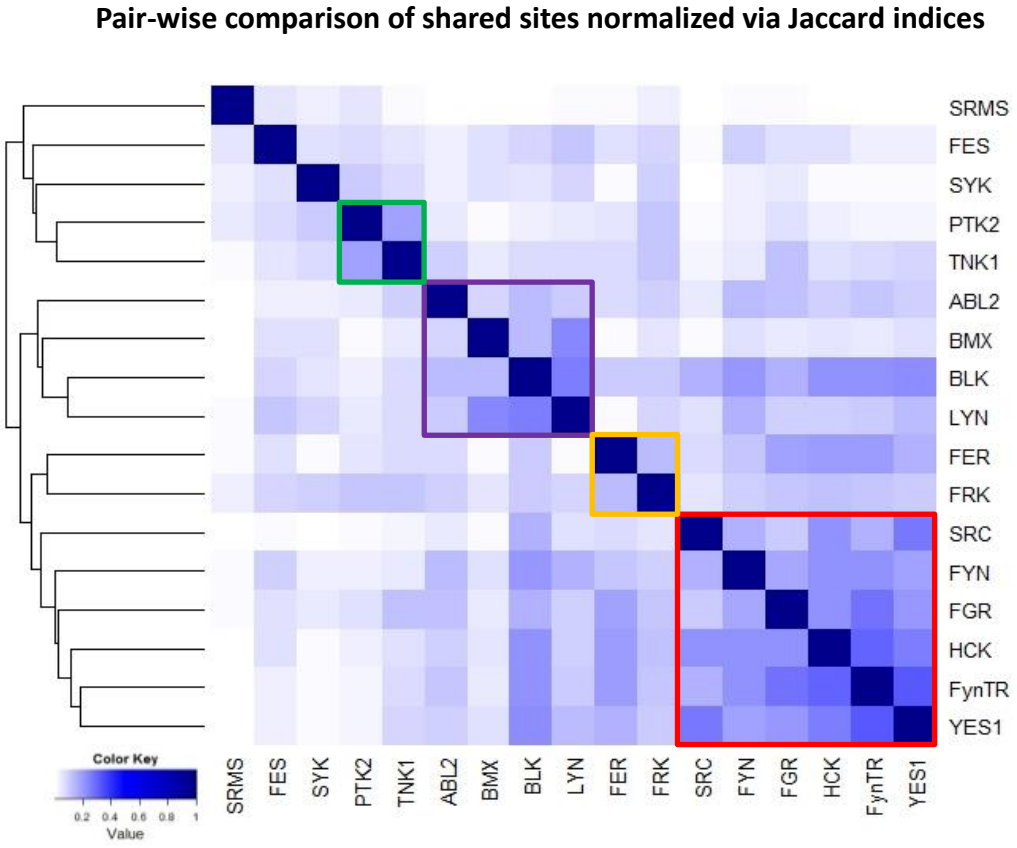


Figure 15: Heatmap showing the specificity overlap between all kinases via Jaccard indices and hierarchical clustering. Colored boxes indicate clusters of NRTKs denoted as appearing to be more similar as compared to all other NRTKs.

Tissue specific expression and subcellular localization, and hence spatial separation, may explain stronger specificity overlap for some kinases (Alexander et al., 2011). Performing hierarchical clustering of NRTK expression in different tissues using RNA sequencing data (Fagerberg et al., 2014) or Expression Sequence Tag (EST) data (Liu et al., 2008) showed that kinases sharing substrate overlap in yeast, i.e. cluster by Jaccard indices, tend to be differentially expressed in human tissues (Figure 16). For instance, PTK2 and TNK1 are most similar with respect to yeast targets and form a cluster framed in green (Figure 15) however, show strong tissue expression dissimilarity as indicated by the dendrogram in Figure 16. Kinases which cluster by tissue expression tend to not cluster by target similarity. For instance, PTK2 (green cluster) and FYN (red cluster) appear to be expressed most similarly in tissues (Figure 16A and B) whereas their yeast substrate overlap is rather small (Figure 15). Moreover, BLK and BMX are less similar to each other than to LYN within the violet target similarity cluster (Figure 15). With respect to tissue specific expression, BLK and BMX are within one major branch of the dendrograms whereas LYN appears to be expressed in different tissues similar to other kinases and hence is placed on another major branch (Figure 16A and B).

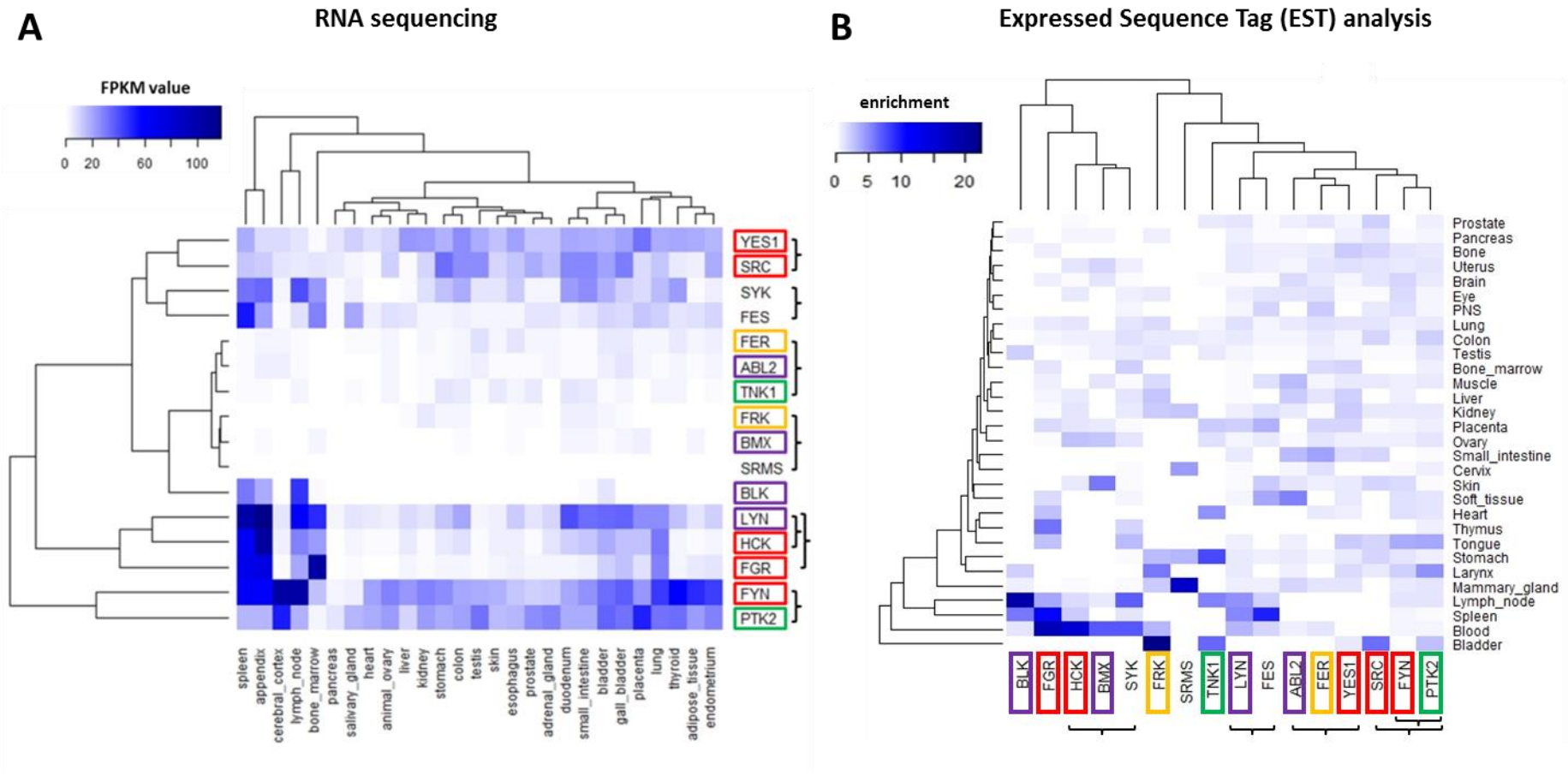


Figure 16: Heatmap depicting tissue specific expression of all NRTKs in this thesis by hierarchical clustering of (A) RNA sequencing data from Fagerberg et al. (2005) and (B) Expression sequence tag (EST) sequencing data from Liu et al. (2008). Colored frames around NRTK names refer to the identified clusters in Figure 15.

3.6. Analysis of protein disorder in yeast phosphorylation sites

An important systematic classification of PTMs and a potential confounding factor is local protein disorder (Woodsmith et al., 2013). A strong preference of locating phosphorylation sites in regions of intrinsic disorder was suggested and that serine and, to a lesser extent, threonine phosphorylation rarely occurs in surface exposed, ordered regions. This feature is distinct from tyrosine phosphorylation (Iakoucheva et al., 2004). Therefore, protein disorder was compared between measured phosphorylated tyrosine residues and all tyrosine residues in the *Saccharomyces cerevisiae* proteome using the tools DISEMBL (Linding et al., 2003), DISOPRED (Ward et al., 2004), and IUPRED (Dosztanyi et al., 2005) as shown in Figure 17. DISEMBL and DISOPRED are both predictors based on machine learning algorithms. Residues are classified using artificial neural networks and the algorithm is trained on high-resolution X-ray crystal structures of proteins where disorder is identified for residues present in sequence records, but with coordinates missing from the electron density map. DISEMBL employs additional definitions of disorder including secondary structure representing coils, loops, and in particular “hot loops” which were determined to have high mobility (Linding et al., 2003). IUPRED is a protein disorder predictor based on estimation of the potential of polypeptides to form inter-residue interactions providing stabilizing energy for globular proteins to overcome the entropy loss during folding (Dosztanyi et al., 2005). Disordered sequences are different and do not show this potential and can hence be distinguished. As expected, tyrosine sites contain only 2.5 (DISEMBL), 8.5 (IUPRED), or 17 (DISOPRED) percent protein disorder compared to 9, 17.5, or 29 percent protein disorder in the genomic background, respectively. Strikingly, the percentage of phosphorylated tyrosine residues from our data mapped to disordered regions is similar to the percentage of all yeast tyrosine residues residing in disordered regions - even when the normalized to spectral counts. This suggests that in yeast tyrosine residues were not overly phosphorylated in disordered and hence sterically less constraint regions, but specifically targeted by the human kinases independent of this characteristic structural property.

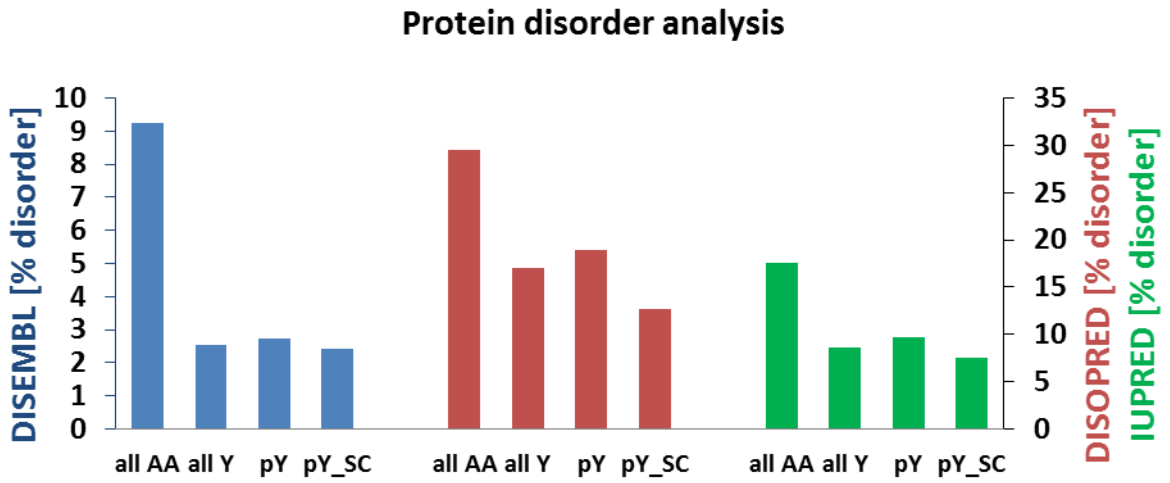


Figure 17: Protein disorder analysis. Yeast disorder predictions by DISEMBL (blue bars, left vertical axis) and DISOPRED (red bars, right vertical axis), and IUPRED (green bars, right vertical axis). The percentage of disorder in respect to the yeast genome, all yeast tyrosine residues, our measured phosphorylated tyrosine residues and summed spectral counts of measured phosphorylation tyrosine residues is shown.

3.7. Phylogenetic analysis of phosphorylation sites from yeast to human

In order to investigate conservation of the phosphorylated yeast tyrosine residues multiple sequence alignments were generated across eukaryotic phylogeny from yeast to human using the INPARANOID database. In summary, the orthologous sequences for 479 yeast ORFs with at least one measured tyrosine phosphorylation site and the presence of a human orthologous sequence were aligned. Tyrosine signaling can be regarded as a hallmark of multi-cellularity. Therefore, sequences of a maximum of 20 species without tyrosine signaling machinery (noTK-group) and up to 17 species, including human, with evolved three tier tyrosine signaling protein sets (TK-group) were compared (Figure 18).

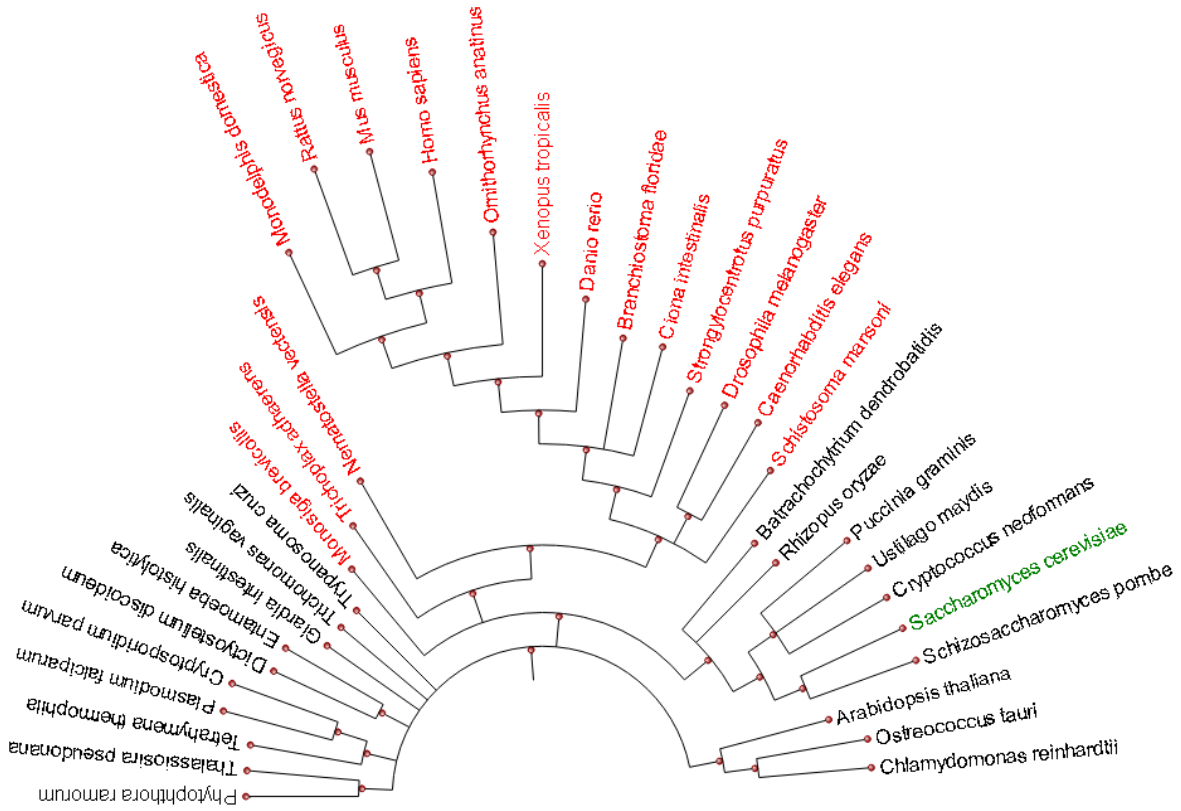


Figure 18: Phylogenetic tree of analyzed species having three tier tyrosine signaling evolved (red) and species not having three tier tyrosine signaling evolved (black). Baker's yeast is labeled green. The phylogenetic tree was generated by retrieving the taxonomy from the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) and using the python environment ETE2.2 (Huerta-Cepas et al., 2010) for visualization.

The alignments covered 942 phosphorylated and 9644 non-phosphorylated tyrosine residues in these proteins. In order to assess the conservation of sites through phylogeny between the TK-group and noTK-group a "Y-score" was calculated as the log2 ratio of the fraction of tyrosine residues at a given position (equation below).

$$Yscore = log2\left(\frac{Ycount\ TKgroup}{number\ of\ TKgroup\ members} \div \frac{Ycount\ noTKgroup}{number\ of\ noTKgroup\ members}\right)$$

Thus, a high Y-score indicates preferential conservation of tyrosine residues in species having evolved tyrosine signaling. Interestingly, 63 of the 310 yeast sites that locally align to a tyrosine in human were reported to be phosphorylated in human as well (Table 4). Figure 19 shows 3 representative alignments where both sites, yeast and human, were measured or reported, respectively.

Symbol Hs	GeneID Hs	Site Hs	Symbol Sc	AccNo Sc	Site Sc	Y-score	Predicted kinases
ACSL4	2182	582	**FAA4**	YMR246W	561	0	FRK
ACTG1	71	53	**ACT1**	YFL039C	53	0.234	ABL2, BMX, PTK2, TNK1
ACTG1	71	166	**ACT1**	YFL039C	166	0	ABL2, BLK, FGR, FYN, HCK, SRC, YES1
ACTG1	71	198	**ACT1**	YFL039C	198	0.515	FER, PTK2, TNK1
ASS1	445	133	**ARG1**	YOL058W	132	0.415	BLK, FGR, HCK, SRC, TNK1, YES1
CDK3	1018	19	**CDC28**	YBR160W	23	0.263	ABL2, BLK, FGR, FYN, HCK, SRC, TNK1, YES1
CKS2	1164	17	**CKS1**	YBR135W	39	1	FRK, SYK
DDX5	1655	202	**DBP2**	YNL112W	221	-0.245	FYN, LYN
EEF1A1	1915	29	**TEF2**	YBR118W	29	0.052	HCK
EEF1A1	1915	254	**TEF2**	YBR118W	252	-0.093	FRK
EEF2	1938	373	**EFT2**	YDR385W	357	0	FER, FGR, FRK, HCK, TNK1, YES1
EIF4A3	9775	54	**FAL1**	YDR021W	39	0.206	BMX
ELP3	55140	329	**ELP3**	YPL086C	338	0	BLK, FER, FGR, HCK, TNK1, YES1
ENO3	2027	131	**ENO1**	YGR254W	131	0.248	FER, FES, FYN, HCK
GAPDH	2597	41	**TDH3**	YGR192C	40	0.29	FER, FGR, FYN, TNK1, YES1
GAPDH	2597	140	**TDH3**	YGR192C	138	0.138	SYK, TNK1
GAPDH	2597	314	**TDH3**	YGR192C	312	0	FGR, FRK, SYK, TNK1
GAPDH	2597	320	**TDH3**	YGR192C	318	0	BLK, BMX, SYK, TNK1
GAPDH	2597	41	**TDH2**	YJR009C	40	0.415	FER, FGR, YES1
GOT1	2805	71	**AAT2**	YLR027C	70	0	ABL2, BLK, FGR, FRK, HCK
HIST1H4	8359	52	**HHF1**	YBR009C	52	0	ABL2, BLK, BMX, FRK, FYN, HCK, LYN, SRC, YES1
HIST1H4	8359	89	**HHF1**	YBR009C	89	0.152	BLK, BMX, LYN
HSP90AB1	3326	56	**HSP82**	YPL240C	47	0.193	HCK
HSPA4	3308	626	**SSE2**	YBR169C	579	0	FYN, YES1
HSPA8	3312	15	**SSA1**	YAL005C	13	0	FGR, FYN,
HSPD1	3329	227	**HSP60**	YLR259C	224	-0.019	FGR, YES1
NOP5/NOP58	51602	342	**NOP58**	YOR310C	343	0.234	ABL2, BLK, FGR, LYN
PCBP1	5093	183	**PBP2**	YBR233W	235	0	TNK1
PDHA1	5160	301	**PDA1**	YER178W	321	0	ABL2
PGAM1	5224	92	**GPM1**	YKL152C	90	0	ABL2, FER, FES, FGR, HCK, TNK1
PGK1	5230	76	**PGK1**	YCR012W	75	0.377	FER, FGR, PTK2
PGK1	5230	196	**PGK1**	YCR012W	194	0.415	FER, FGR

PKM2	5315	83	**CDC19**	YAL038W	59	0.434	FER, FES, FGR, TNK1
PKM2	5315	466	**CDC19**	YAL038W	436	1	PTK2, SYK
PSMA2	5683	24	**PRE8**	YML092C	23	0.788	FER
PSMA2	5683	76	**PRE8**	YML092C	75	0.16	ABL2, BLK, BMX, FGR, FYN, HCK, LYN, SRC, TNK1, YES1
PSMA7	5688	153	**PRE6**	YOL038W	156	0.582	FER, FGR, FYN, HCK, YES1
PSMB3	5691	103	**PUP3**	YER094C	103	0.234	FGR, PTK2
PSMC4	5704	205	**RPT3**	YDR394W	212	0.052	BLK, FER, HCK, PTK2
RAB1A	5861	8	**YPT1**	YFL038C	5	-0.093	FER
RAB8B	51762	5	**SEC4**	YFL005W	17	0	FER, FRK, SRMS
RPL15	6138	62	**RPL15A**	YLR029C	62	0.322	BMX
RPL21	730559	13	**RPL21A**	YBR191W	13	1.65	TNK1
RPL23	9349	38	**RPL23A**	YBL087C	35	0.322	FGR, LYN, TNK1
RPL27A	6157	52	**RPL28**	YGL103W	52	0.078	BLK, TNK1
RPL34	6164	13	**RPL34A**	YER056C-A	13	0.074	FGR
RPL38	6169	43	**RPL38**	YLR325C	52	0	FES
RPL4	6124	52	**RPL4A**	YBR031W	50	0.229	FGR
RPL7A	6130	60	**RPL8A**	YHL033C	55	0.248	ABL2, BLK, FES, FGR, FRK,
RPL8	6132	133	**RPL2A**	YFR031C-A	133	0.067	BMX, FES, FRK, TNK1
RPS11	6205	55	**RPS11A**	YDR025W	53	0	BLK, BMX, FES, FGR, FRK, LYN, PTK2, SYK, TNK1
RPS27	6232	31	**RPS27A**	YKL156W	31	0.971	SYK, TNK1
RPS4X	6191	54	**RPS4A**	YJR145C	54	0.078	BLK, FYN
RPS9	6203	13	**RPS9A**	YPL081W	12	0.604	ABL2, FGR, FYN, LYN, SYK, TNK1, YES1
RPSA	3921	139	**RPS0A**	YGR214W	138	0.29	TNK1, YES1
SGTA	6449	158	**SGT2**	YOR007C	169	1.585	HCK, TNK1
TARS	6897	298	**THS1**	YIL078W	305	0.152	FGR
TKT	7086	275	**TKL1**	YPR074C	294	-0.492	HCK
TUBA4A	7277	399	**TUB1**	YML085C	400	0.152	BLK
VARS	7407	469	**VAS1**	YGR094W	315	0.148	ABL2
VCP	7415	644	**CDC48**	YDL126C	654	0.152	ABL2, BMX
VTA1	51534	285	**VTA1**	YLR181C	310	1	FER, FES
WRNIP1	56897	534	**MGS1**	YNL218W	432	0	BMX

Table 4: Homology predicted human pY-sites by sequence alignments (INPARANOID database). Hs = *H. sapiens* and Sc = *S. cerevisiae*. Validated kinase-substrate relationships indicated by red letters. Proteins involved in glucose metabolism are marked with red background color.

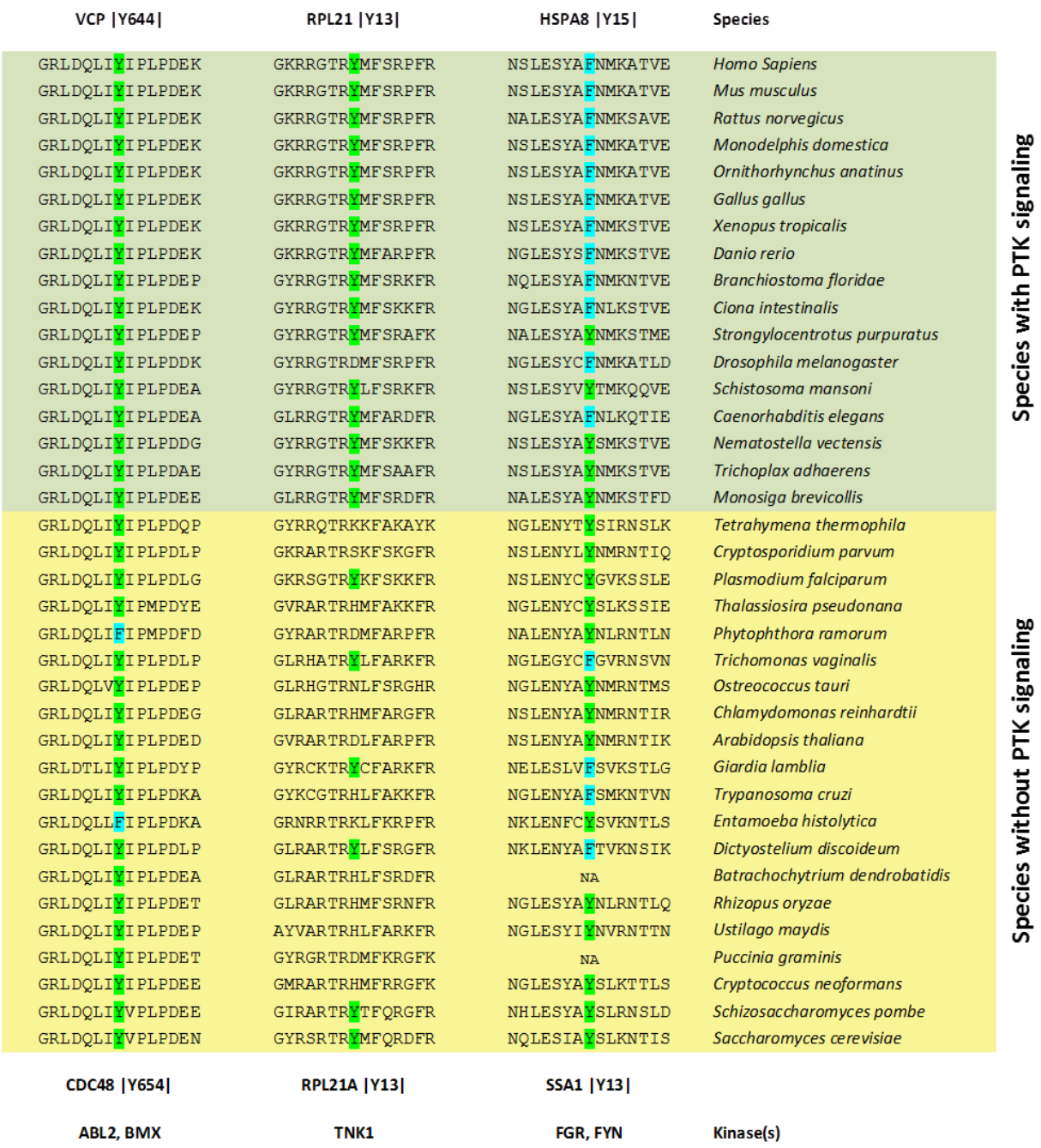


Figure 19: Three selected examples for multiple sequence alignments. Below the alignments the yeast protein identifier is provided together with the measured phosphorylation site. Above the alignments the human ortholog identifier and corresponding site is shown. Species are grouped into having and having not evolved tyrosine signaling.

A Y-score close to zero (0.152) for Y654 in cdc48/VCP, for example, suggested an even distribution of tyrosine between the two groups. Indeed, the alignment showed strong conservation of the phosphorylation site including the surrounding amino acids across all species suggesting that the tyrosine residue was structurally necessary, but also that kinases targeting Y654 in yeast may modify the corresponding site in human. As shown in Figure 20, the 942 measured and aligned phosphorylated yeast tyrosine residues are overall significantly more conserved (**all YpY**, 26.9%;

p=0.0008, Figure 20) within the TK-group than all other tyrosine in yeast (**all YY**, 22.3%, Figure 20). Moreover, in two third of 310 cases yeast pY-sites locally aligned to a tyrosine residue in human orthologs. These tyrosine residues were significantly higher conserved in the TK-group of species (**YpY-HY**; 62.3%, p=0.0044, Figure 20) in comparison to all other yeast tyrosine residues aligning to a human tyrosine residue (**YY-HY**, 54.4%; Figure 20). For example, a high Y-score of 1.65 for Y13 in RPL21 phosphorylated by TNK1 represents this situation (Figure 19). Only six tyrosine residues aligned in the noTK-group while almost full conservation of tyrosine residues within the TK-group of species hints towards a possible functional relevance for this phospho-site. In a nutshell, all tyrosine residues targeted in yeast and pY-sites with an orthologous tyrosine residue in human were better conserved in species having active PTKs and a median Y-score larger than zero was observed (**YpY-HY**; Figure 21). This supports the notion that a subset of the pY-sites measured in yeast may be transferred to human and may provide direct evidence for the kinases involved.

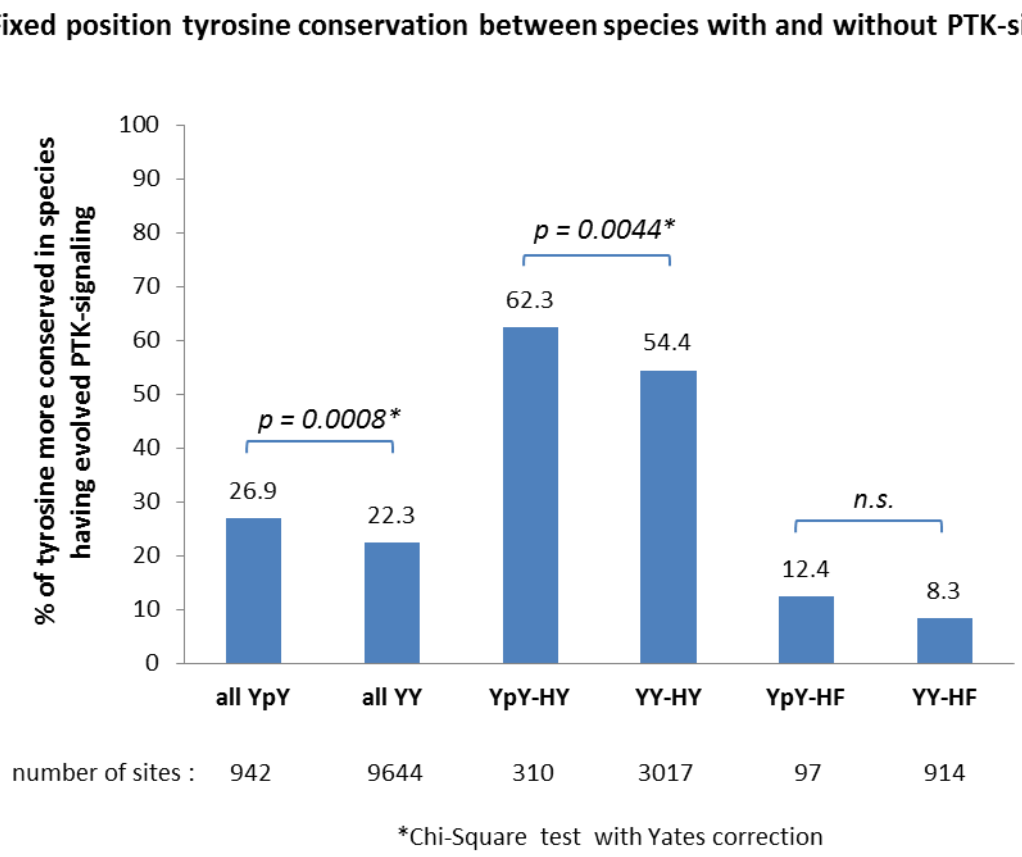


Figure 20: Fixed position tyrosine conservation between species with and without evolved NRTK-signaling. Comparison of tyrosine conservation of all phosphorylated tyrosine in yeast (all YpY) with all tyrosine in yeast (all YY), of phosphorylated tyrosine in yeast conserved in human (YpY-HY) with all tyrosine in yeast conserved in human (YY-HY) and of all phosphorylated tyrosine in yeast counter-selected in human (YpY-HF) with all tyrosine in yeast counter-selected in human (YY-HF). n.s. = non-significant.

Tan et al. observed negative correlation of total protein tyrosine content of organisms with increasing number of cell types or increasing number of predicted PTKs from *S. cerevisiae* to *H. sapiens* (Tan et al., 2009b). In addition, a significantly smaller fraction of amino acids in human proteins are tyrosine than in their one-to-one yeast orthologs (Tan et al., 2009b). It was suggested that one reason for this selection could be a beneficial reduction of tyrosine phosphorylation. Therefore, it was examined whether measured pY-sites in yeast are substituted in human to the structurally most similar, not modifiable amino acid phenylalanine (F). The fraction of conserved phenylalanine residues within the TK-group was almost identical (**YF-HF**; 54.2%, not shown) to the fraction of conserved tyrosine residues (**YY-HY**; 54.4%, Figure 20). However, 97 of the phosphorylated tyrosine residues in yeast were phenylalanine in human orthologs. For these sites a low fraction of tyrosine residues (**YpY-HF**; 12.4%, Figure 20) was found within the TK-group with a median Y-score below zero (Figure 21). The percentage of tyrosine residues within the TK-group in respect to all yeast tyrosine residues with a phenylalanine conversion in human was not significantly lower (**YY-HF**; 8.3%, Figure 20). Additionally, for more than half of yeast pY-sites showing phenylalanine substitution in human the noTK-group had a slightly higher fraction of tyrosine residues (**YpY-HF**; 53%, not shown). One can hypothesize that those sites may be phosphorylated by human PTKs in yeast but were selected against in human and the TK-group. In the example shown in Figure 19, Y13 on ssa1/HSPA8 targeted by FYN and FGR is largely replaced by phenylalanine within the TK-group resulting in a negative Y-score of -1.4. Both Y to F and F to Y conversion from yeast to human showed a similar distribution between species having and having not tyrosine signaling evolved when focusing on proteins that have been phosphorylated in the yeast experiments; however, with a median Y-score and F-score close to zero, respectively (Figure 21).

In summary, 63 of 310 pY-sites measured in yeast with a human orthologous tyrosine were also reported to be phosphorylated in human (Table 4) and those tyrosine residues were preferentially found in species having evolved tyrosine signaling (**YpY-HY**; Figure 20). Moreover, 97 measured yeast pY-sites were phenylalanine residues in human with a negative Y-score (a relative higher frequency of tyrosine residues at corresponding positions in species which did not evolve PTKs). This result hints towards cases of potential counter selection of tyrosine residues in metazoans as suggested by Tan et al. (2009b).

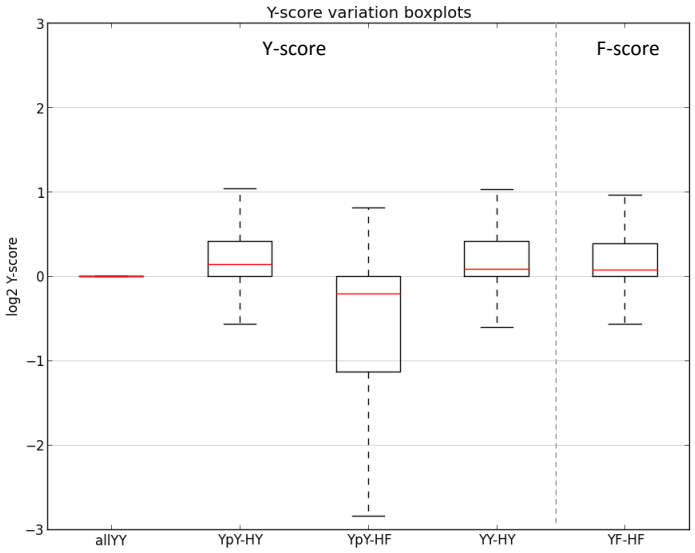


Figure 21: Boxplots of Y-score and F-score variation for each of the different alignment cases: all yeast tyrosine residues (allYY), measured yeast pY-sites with human orthologous tyrosine residue (YpY-HY), measured yeast pY-sites with human orthologous phenylalanine residue (YpY-HF), all yeast tyrosine residues with human orthologous tyrosine residue (YY-HY), and all yeast phenylalanine residues with human orthologous phenylalanine (YF-HF, F-Score).

Querying all 63 homology predicted pY-sites (representing 53 proteins) for enriched gene ontology (GO) terms using a tool developed at Stanford University (Boyle et al., 2004) and implemented at Princeton University yielded a total of 168 partly overlapping, enriched GO-terms with many terms related to cellular metabolism and the translational machinery. A reduced list of enriched GO-terms is visualized in Figure 22 and clusters of GO-terms related to the two main umbrella terms metabolism and translation are marked by circles.

Strikingly, among the enriched metabolism related GO terms was the term “glycolytic process” (p-value 0.0000122) as indicated in Figure 22 (also at axis labels). The GO-term enrichment involves six homology predicted targets, approximately 11 percent of all predicted orthologs namely pyruvate enolase 3 (beta, muscle) (ENO3), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), dehydrogenase alpha 1 (PDHA1), phosphoglycerate mutase 1 (PGAM1), phosphoglycerate kinase 1 (PGK1), and pyruvate kinase muscle (PKM). Those homology-predicted glycolytic enzymes together with transketolase (TKT) and cytoplasmic aspartate aminotransferase (GOT1) were also included in enriched GO terms “monosaccharide metabolic process” (p-value 0.0000215) and “generation of precursor metabolites and energy” (excluding GOT1; p-value 0.00418). Indeed, the six homology predicted targets involved in glycolysis almost represent the entire glycolytic pathway as depicted in a schematic representation of glycolysis in the discussion section (Figure 50). In summary, highly conserved cellular processes were targeted by expressed human NRTKs in yeast cells and hence it may be feasible to directly infer KSRs by mapping modification sites to one-to-one human orthologs.

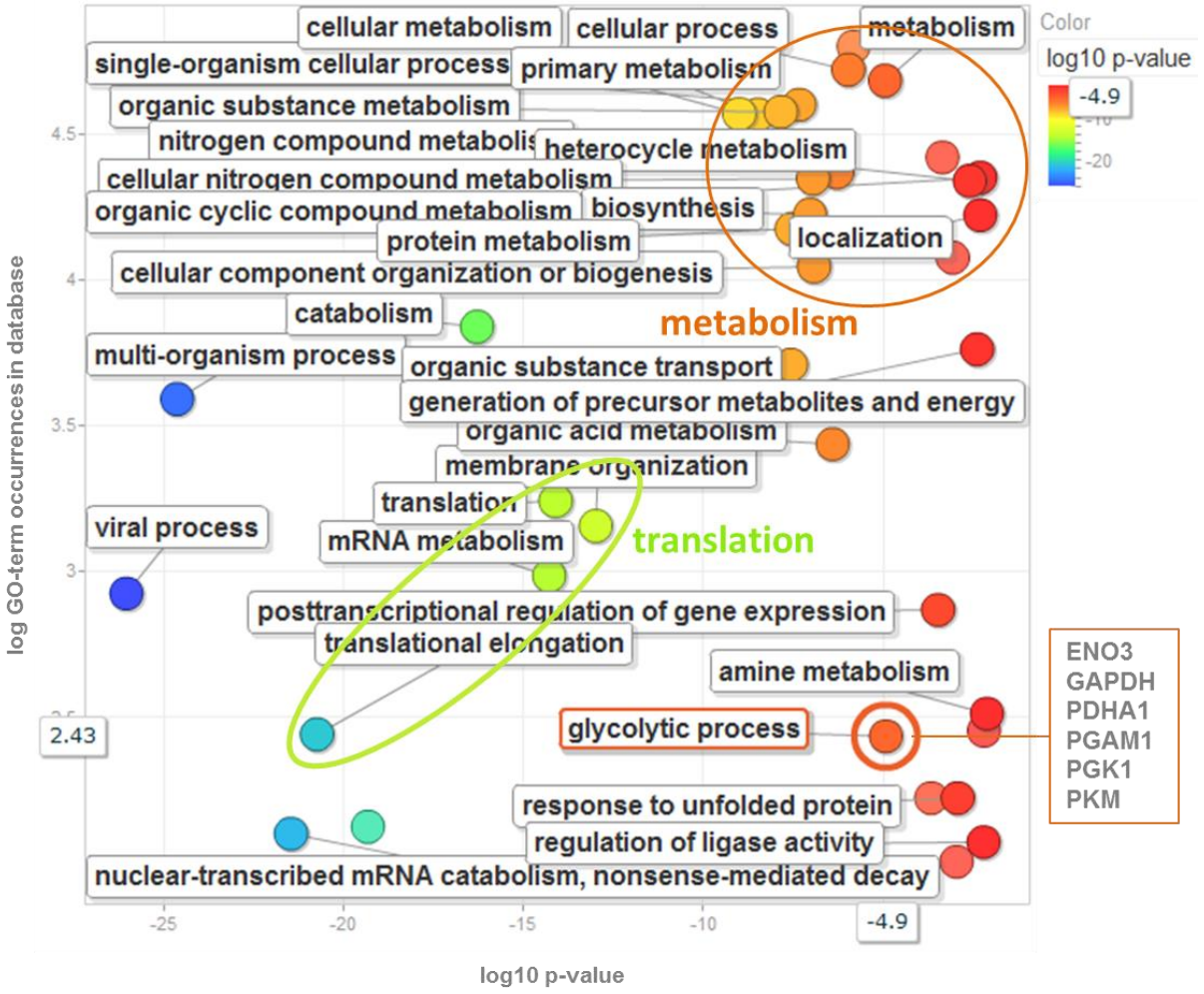


Figure 22: Visualization of process gene ontology (GO) term enrichment using REVIGO (Supek et al., 2011) and a reduced output list from GO::TermFinder (Boyle et al., 2004). The log of GO-term frequency within the GO database GOA homo sapiens is plotted against log10 p-values (also indicated by dot color). Approximate clusters of GO-term enrichments corresponding to two umbrella terms are encircled and the enriched GO term "glycolytic process" is highlighted including axis labels and associated NRTK substrates.

3.8. Determination of linear sequence motifs

Linear amino acid sequence motifs are considered as one of the main substrate properties for kinase recognition and consensus motifs were previously generated by linear peptide screening (Mok et al., 2010, Miller et al., 2008). Thus, sequence motifs were generated by estimating the frequencies of amino acids framing the modification sites within each of the measured 17 kinase target sets in yeast. In contrast to previous methods, the generated large yeast substrate sets contain targeted sites from phosphorylation of fully folded yeast proteins in a highly-crowded cellular environment which is a very distinct representation of a potential kinase substrate space compared with randomized peptide arrays. Using iceLogo (Colaert et al., 2009) kinase specific sequence were generated for seven amino acids up- and downstream of the modification sites for all tested NRTKs

(Appendix 6.1). Linear sequence motifs are typically assessed from alignment of 11mer or 15mer centered at an amino acid representing a modification or binding site. Several motif prediction tools are implemented (Crooks et al., 2004, Schwartz and Gygi, 2005, Amanchy et al., 2007, Schwartz et al., 2009, Ritz et al., 2009, O'Shea et al., 2013) including iceLogo (Colaert et al., 2009) which is a probability based method employing a PSSM generated by comparison between amino acid frequencies within positive and negative modification sequence sets.

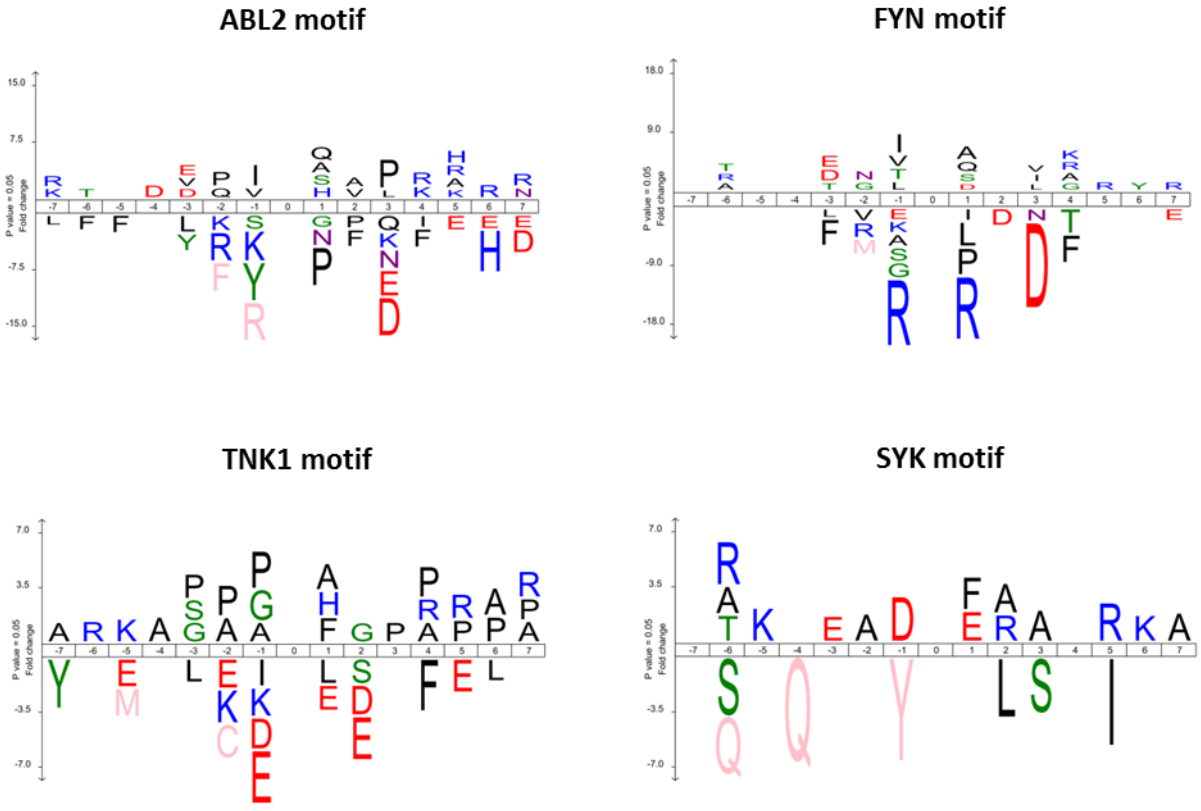


Figure 23: Amino acid sequence motifs for ABL2, FYN, TNK1 and SYK generated from the respective yeast target sets.

In contrast to the widely used motif-x algorithm (Chou and Schwartz, 2011), iceLogo also determines under-represented amino acid frequencies and fold-chance enrichment values within quasi position weight matrices (PWM) can be obtained for each kinase. In order to retrieve amino acid enrichment values for motif generation a genomic background set has to be defined. Therefore, a “yeast tyrosine proteome” reference set was created by retrieving all tyrosine residues in the entire yeast proteome (13577 identifiers; downloaded 01/06/2011 from SGD) as 15mers with seven amino acids flanking each tyrosine herein after called Y-sites (95819 Y-sites) and a second reference set was created containing all Y-sites of the “expressed yeast proteome” (17224 Y-sites), i.e. all proteins measured over all experiments. From both sets all measured 1433 pY-sites were removed. Motif performance

analysis showed minor differences in retrieving the input from the proteome using either of the two reference sets however, the “expressed yeast proteome” reference set was chosen for motif generation to ensure expression and presence of those proteins. Example motifs (the pY is at position 0) are shown for ABL2, FYN, TNK1, and SYK kinase in Figure 23. Additionally, the full data is represented by iceLogo enrichment plots which is not the case for frequency plots retrieved by motif-x. Of note, only a fraction of the measured pY-sites yielded a significant score when assessed with the motifs (see motif refinement, paragraph 3.11).

Each set of pY-sites resulted in an informative motif containing differing characteristics such as the enrichment of single amino acids at certain positions relative to phosphorylated tyrosine or stretches of biochemically related amino acids flanking the modification site. For example, positively charged amino acids C-terminal to the modification site for ABL2, FYN, and TNK1 or an enrichment of proline spanning the entire motif of TNK1 can be recognized from respective motif representations (Figure 23). There are many other motif features which can be detected by comparing motifs between kinases and hence differences and similarities in amino acid preference may be detected.In order to provide a better overview of the whole data set motif features were extracted and grouped (Figure 24). Feature extraction was based on manual comparison of amino acid occurrences between all generated sequence motifs. Extracted features were grouped if they were observed to co-occur within the same motifs. If features co-occurred only in a subset of motifs, the combination was introduced as another motif feature for the subset of NRTKs. However, grouping of motif features was not conducted if individual features are common for related or families of NRTKs or may explain NRTK substrate specificity overlap as observed in yeast. Of note, the features may not indicate inter-positional dependencies which were suggested to be very small within sequence motifs of S/T kinases by Joughin et al. (2012) on a statistical basis. The authors cross-validated performance of probabilistic first order models to second order models by scoring three published substrate datasets. The authors concluded that addition of second order, inter-positional information does not improve or may even lower the quality of kinase substrate predictions. The most prominent feature within almost all motifs are aliphatic amino acids (V/L/I) on the position -1 N-terminally to the phosphorylated tyrosine residue as previously reported for SFKs and related NRTKs including ABL kinases (Songyang and Cantley, 1995, Colicelli, 2010, Deng et al., 2014). Of note, leucine in this position is also the preferred contextual residue in pY-recognition by the pY-antibodies used here whereas charged residues in this position are hardly accepted (Tinti et al., 2012). Only TNK1, PTK2, SYK, and SRMS motifs do not contain this feature. These four NRTKs also share little target similarity overlap with other kinases (Figure 15).

Notably, all retrieved iceLogo motifs harbor a strong underrepresentation of some amino acids at certain positions suggesting a mechanism within yeast substrate sequence motifs which may prevent (aberrant) kinase targeting. Substrate sequences may contain “non-permissive” specific amino acids that oppose a kinase-substrate interactions by steric hindrance or charge-based repulsion and which hence prohibit a specific residue, or class of residues, at one or more positions in putative kinase substrates (Liu et al., 2010). By using fold change as the height of an over-represented amino acid in the iceLogo application, the height of under-represented amino acids is comparable. Among under-represented amino acids, there are some colored pink which indicates that these residues never occur within the positive set and are hence significantly under-represented compared to the background sequences in yeast. The fold-change for these instances is infinite due to a division of the number of occurrences within the background set by zero instances measured. However, if more residues are regulated having finite fold-change, the height of the infinite amino acids will be displayed 10 percent larger than the height of the largest finite amino acid. In all SYK substrates a tyrosine residue (Y) was never present at position -1 while a glutamine residue (Q) never appeared at position -4 and -6 (Figure 23). In contrast, there are a few instances in the SYK substrate set where an isoleucine was measured in position +5 resulting with a finite, still significant fold-change and hence the letter size was limited. Apart from SRMS, there is at least one amino acid at at least one position never present in the target set sequences and hence colored pink in each kinase-specific linear sequence motif.

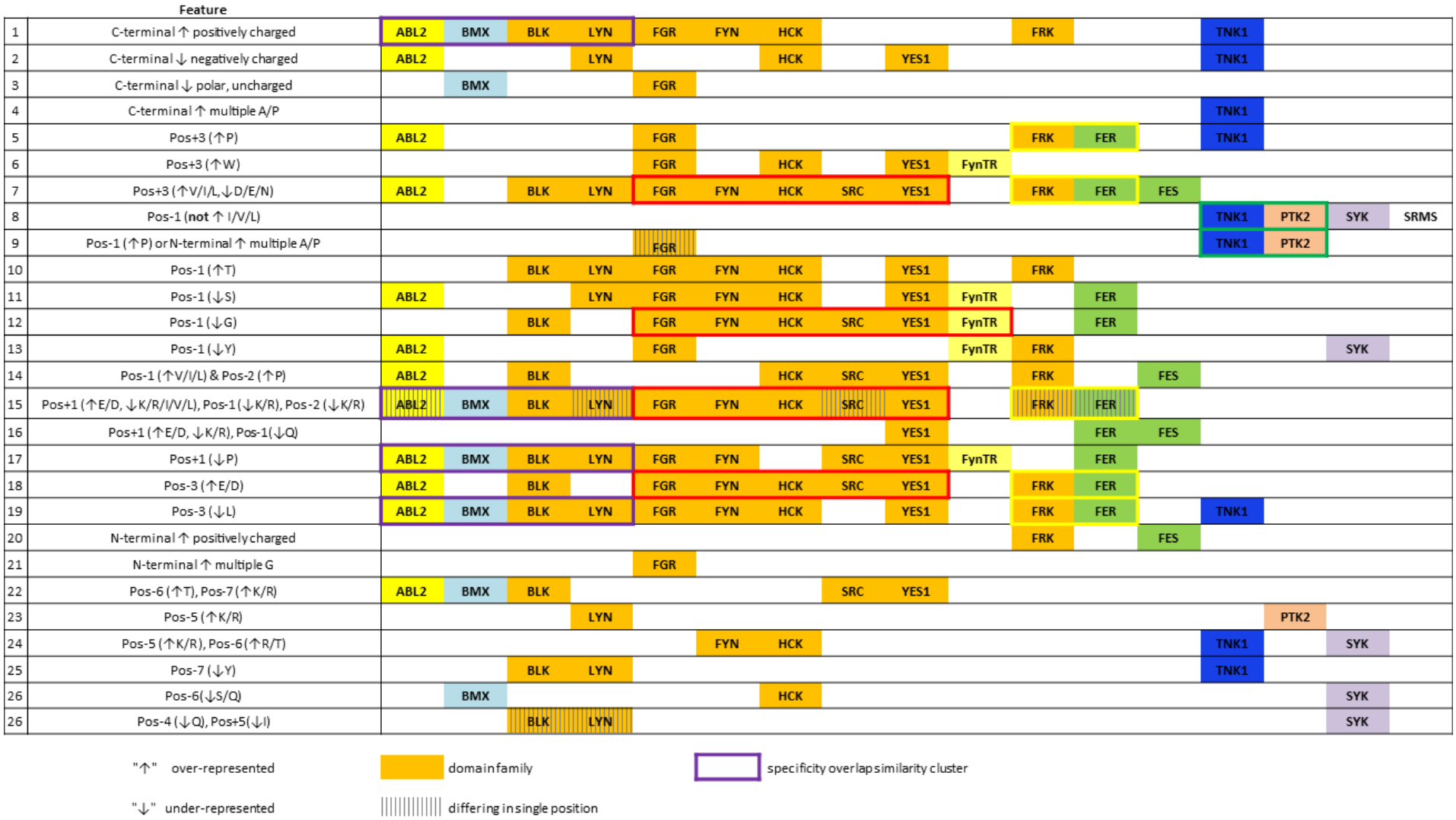


Figure 24: Extracted sequence motif features. Kinases are arranged according to target specificity similarity clustering as shown in Figure 15 while background colors indicate kinase families according to domain structure. A colored box (as in Figure 15) indicates that all members of a cluster have the feature.

Comparing motif features to target similarity clustering could unravel similar sequence targeting mechanisms for kinases within clusters. For instance, ABL2, BMX, BLK, and LYN share many yeast substrates and all four kinases show a preference for modification site C-terminal stretch of positively charged amino acids. Interestingly, ABL2 and LYN disfavor negatively charged amino acids in this stretch while BMX disfavors polar, uncharged residues. Moreover, all four kinases oppose proline at position +1 and leucine at position -3 (Figure 24, feature 17 and 19) and share feature 15 with the exception of differing ABL2 preferences for position +1 and that LYN appears to tolerate positively charged residues at position -2. These motif properties are not unique for the four kinases, but may explain in part the clustering in the analysis of specificity overlap. Only few of the motif features are shared among all those four cluster members and it appears that specificity for each kinase is provided by variable combinations of motif features. FRK and FER were also clustered by targeting similar yeast proteins, but share many targets with SFKs. The motifs for FER, FRK, and SFKs are accordingly similar as four features (7, 12, 15, 18) are shared. Only the discovered feature 5, a preference for proline at position +3, shared between FER and FRK could partly explain the difference between the two similarity clusters. Interestingly, Feature 5 is best described for the ABL kinase family motif (Colicelli, 2010), however; the analysis here shows that also other kinases employ this targeting property. TNK1 shared 87 of its 385 pY-sites with PTK2 covering 71 percent of 122 sites, and both kinases showed only little overlap to all other NRTKs (Figure 15). Moreover, their motifs are different to all other NRTK motifs generated by showing preference for alanine and proline N-terminal to the tyrosine (Feature 9, Figure 24). There are further motif characteristics unique to other kinases. Only FGR favors a stretch of glycine N-terminally and SYK is the only kinase with a strongly disfavoring glutamine at position -4 and -6 while only FES disfavors glutamine at position -2, for example. Due to high number of motif features shared among kinases it appears that the specific combination of these features renders a sequence motif unique what agrees with step-wise evolution of kinase domains. In summary, patterns of positive and negative enrichment characteristics were detected by comparison of sequence motifs between NRTKs. There are kinase-specific motif features and features shared among the kinases analyzed which are embedded in kinase-specific combinations.

3.9.Estimating linear sequence motif performance

The performance of a linear sequence motif is the combination of sensitivity and specificity when identifying kinase substrate sites based on their primary amino acid sequence (Baldi et al., 2000, Miller et al., 2008). Using specific PMWs retrieved from iceLogo for each kinase for scoring the "yeast tyrosine proteome" (= search space) and assigning the truly targeted Y-sites (true positive set) using a binary identifier (1 for positive, 0 for negative) Receiver-Operating-Characteristic (ROC) curve analyses were performed in order to investigate how well the kinase specific pY-sites are represented by the kinase specific "motif score". An ROC curve describes the False-Positive-Rate (FPR) as function of the True-Positive-Rate (TPR) by going through a sorted list of all scored sites starting with the highest score and varying threshold. The x-axis describes the specificity (TRP) and the y-axis the selectivity (1-FPR) for the classifier. The steeper the curve rises at high TPR values, the more true positives have highest scores assigned at a given cut-off. Performance can be expressed by the "area-under-curve" (AUC) value whereas an AUC value of 1 would mean all true positives have highest scores assigned (ideal performance) while an AUC value of 0.5 would mean the classifier, here the "motif score", determines positives at random (worst performance). ROC curves were generated using the ROCR-software (R-project.org) using a scored reference proteome as search space. Because the motif is intended to classify a set of 13240 human phosphorylation sites, an equal number of randomly picked Y-sites was retreived from the "yeast tyrosine proteome" comprising the reference set. In the main text of this thesis, the linear sequence motif analysis is exemplarily demonstrated on the example of ABL2 kinase. As it can be seen in Figure 25 the ROC curve for the ABL2 motif in retrieving ABL2 targets shows a AUC value of approximately 0.9 (A, colored curve) and, for comparison, testing the ABL2 motif for retrieving the exclusive targets of all other kinases combined results in an AUC value of 0.67 (A, black curve).

A strong drop in performance testing the motif on all other kinase target sets suggests the motif is kinase specific and may be used for target prediction. As the negative set is very large compared to the number of kinase targeted sites and the positive set and training set were identical, the AUCs of kinase varying between 0.82 of 0.9 are assumed to be an overestimation of the actual motif performance. In order to provide an alternative performance test tenfold cross-validation was performed with all generated sequence motifs (Appendix 6.2). Due to the limited number of reported kinase-substrate relationships in public databases for the majority of NRTKs, it was not possible to retrieve sufficiently large independent positive sets for systematic motif performance testing. Therefore, cross-validation was performed where ten percent of the kinase target sets were randomly removed and a new sequence motif generated using the remaining 90 percent of hits for each kinase. The reference set was subsequently scored applying the new motif and binaries

assigned labeling the omitted, independent ten percent of targeted pY-sites. This random generation of an independent positive set was performed 100 times and the average performance of the motif determined (Figure 25B, black curve). The AUC between kinases varied from 0.67 to 0.77 whereas performance was more robust if more sites were targeted, for example in the case of TNK1 (358 pY-sites) and FGR (454 pY-sites). SRMS had too few targets (28 pY-sites) to perform cross-validation leading to an average AUC of 0.61 with large variation between single randomizations. In summary, linear sequence motifs could be generated for all kinases assayed. Importantly, for most tested kinases no linear sequence motif has been reported to date. For some kinases the motifs recapitulate consensus data (see discussion section) and extends the knowledge by determination of substrate binding prohibitory factors, for instance.

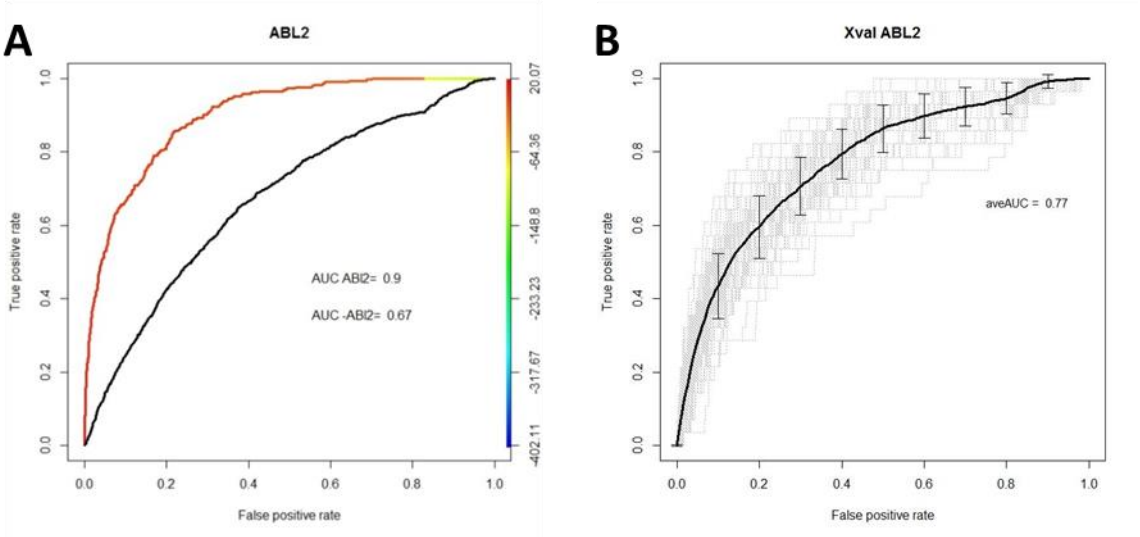


Figure 25: Motif performance analysis (ROC) example. All yeast tyrosine sites were scored using the ABL2 motif and the identification of all measured ABL2 pY-sites (A, colored curve) was compared to finding all other measured pY-sites from all other kinases (A, black curve). Cross-validation was performed by 100 times creating a motif on randomly chosen 90 % of pY-sites in the kinase target set and identification of the remaining 10 % of pY-sites in the scored yeast proteome (B, grey curves). The average over all 100 randomizations is presented including the standard deviation (B, black curve) and average AUC.

3.10. Improvements on sequence motif performance

Motif scoring of phosphorylation sites resulted in kinase dependent, differing range of scores. For example, motif scores ranged from -910.69 to 31.39 for ABL2 and from -1298.98 to 17.43 for FES (Table 5). Hence, scores were converted to accuracy, which is defined as the sum of true-positives and true-negatives divided by the sum of all positives and all negatives, in order to enable inter-kinase comparisons. An optimal accuracy cut-off was to be chosen for minimizing the False-Discovery-Rate (FDR), i.e. the fraction of false discoveries among all discoveries. As there are only limited sets of true positives for only a subset of kinases in human available, it was tested in yeast

how many truly targeted sites by each NRTK can be predicted correctly using varying motif score accuracy cut-offs. Thus, the measured 1433 pY-sites in yeast were scored with all kinases separately and labeled true positive if the site was targeted yielding a list of (16 x 1433 =) 22928 labeled kinase site pairs. Subsequently, the scores were normalized to accuracy values using the ROCR software. Using accuracy cut-offs ranging from 0.94 to 0.9975 in 0.0025 steps the resulting number of true-positive and false-positive kinase site pairs was determined. The more stringent the cut-off was chosen the less targets were predicted, however; with decreasing FDR. Hence, a cut-off was chosen in order to yield as many sites as possible with smallest FDR possible.

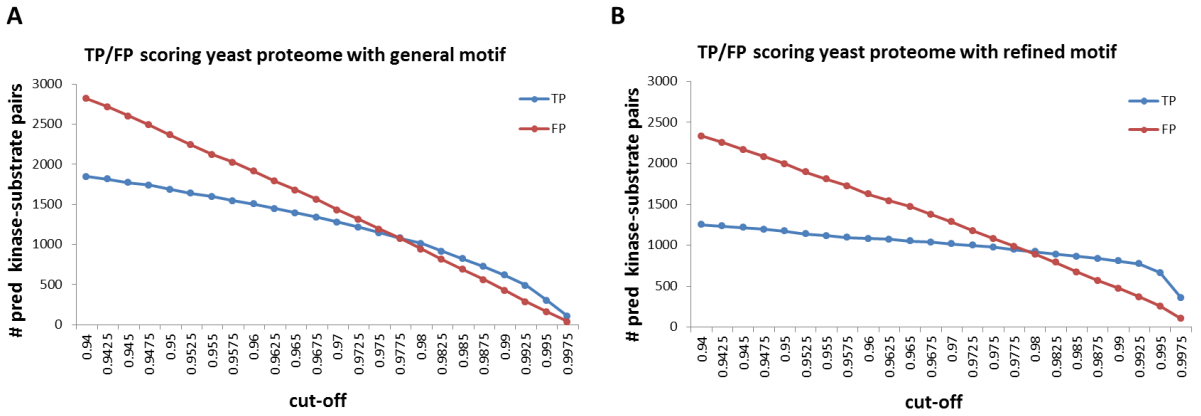


Figure 26: Number of true positives and false positives within predicted kinase-substrate pairs over increasing accuracy cut-off for the original motifs (A) and the refined motifs (B) when scoring all measured 1433 yeast pY-sites with all NRTK motifs.

The left graph in Figure 26 shows that starting at a accuracy cutoff of 0.98 more true positives than false positives were predicted yielding 1010 true positive kinase site pairs of 1950 predicted (51.8%, FDR=0.48). The most stringent cut-off at 0.9975 accuracy yielded 101 true positives of 136 overall predicted targets (74.3%, FDR=0.257). As the number of sites using the maximum cut-off is very small, the ideal cut-off here may be 0.995 predicting 299 (65%) of 456 sites correctly (FDR=0.344). However, the discriminative power of the motif seemed to be relatively small and there was need for improvement. The motifs were generated using all targeted yeast tyrosine. However, as sequence motifs may explain only 20-30 percent of kinase substrate specificity (Linding et al., 2007), only 20 to 30 percent of the targeted sites in yeast may carry motif information. Hence, it was reasoned to improve the specificity of the motif by only taking the 20 percent of best scoring targeted sites for each kinase and redraw a sequence motif, herein called “refined motif” for all NRTKs (Appendix 6.3). Interestingly, the underrepresentation of specific amino acids within the refined motifs was strongly emphasized as observed for ABL2, FYN, and TNK1 in Figure 27. For example, motif feature 19 (Figure 24), an under-represented leucine three residues N-terminal from the modified tyrosine, was detected for 11 kinase motifs. The refined motifs suggested this leucine never appears in the top 20

percent of motif scoring target sites of ABL2, FER, FRK, FYN, HCK, and YES and rarely in the top scoring FGR targets while the feature disappeared in the remaining four refined motifs (Appendix 6.3). Furthermore, it can be observed that leucine (e.g. position +1) and serine (e.g. position -6) are even more strongly disfavored. As a last example, the C-terminal stretch of negatively charged amino acids in the TNK1 motif is more under-represented in the refined motif (Figure 27).

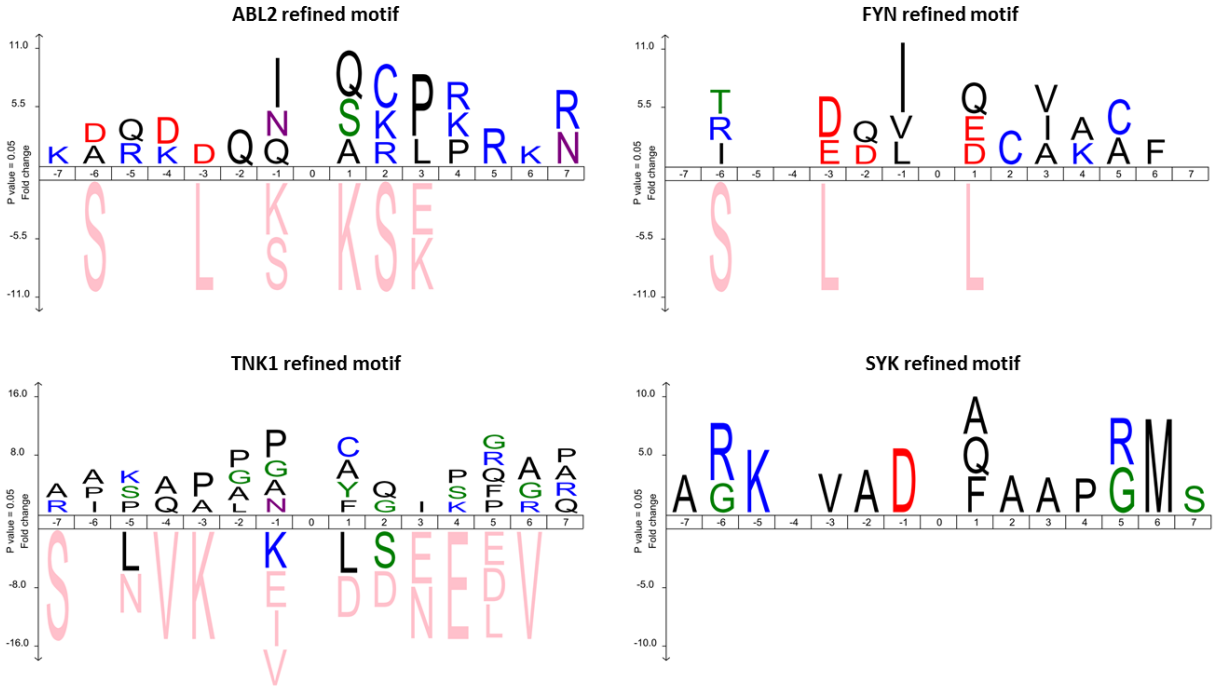


Figure 27: „Refined" amino acid sequence motifs for ABL2, FYN, TNK1 and SYK generated from the 20% of beast scoring sites within the respective yeast target sets.

Performance of the refined motifs was compared to the original motifs by ROC analysis for all NRTKs (Appendix 6.4). As exemplified for ABL2 kinase the refined motif performed worse in finding the all kinase specific targets (Figure 28, red curve) with a drop in AUC from 0.89 to 0.78 however, performed better in assigning the 2 percent top scoring sites (Figure 28, enlargement). The analysis of all kinase site pairs in order to find an optimum accuracy cut-off was repeated using the refined motif. As shown in the right graph in Figure 26 the discrimination between true-positives (here 20 percent best scoring sites per kinase) and false positive was improved for high accuracy cut-offs. While the highest examined cut-off still yielded 355 true-positive of 458 predicted sites (77.5 %, FDR=0.225), the chosen optimum accuracy cut-off of 0.995 yielded 663 true-positives of 918 pY-sites (72.2 %, FDR=0.278) - a clear improvement in comparison to original motif (FDR=0.344) using the same cut-off. This outcome suggests that by using the refined motif and an ideal, very high accuracy cutoff to score and predict human pY-sites would yield the optimal number of predictions while maintaining a very high true positive rate of approximately 80 percent.

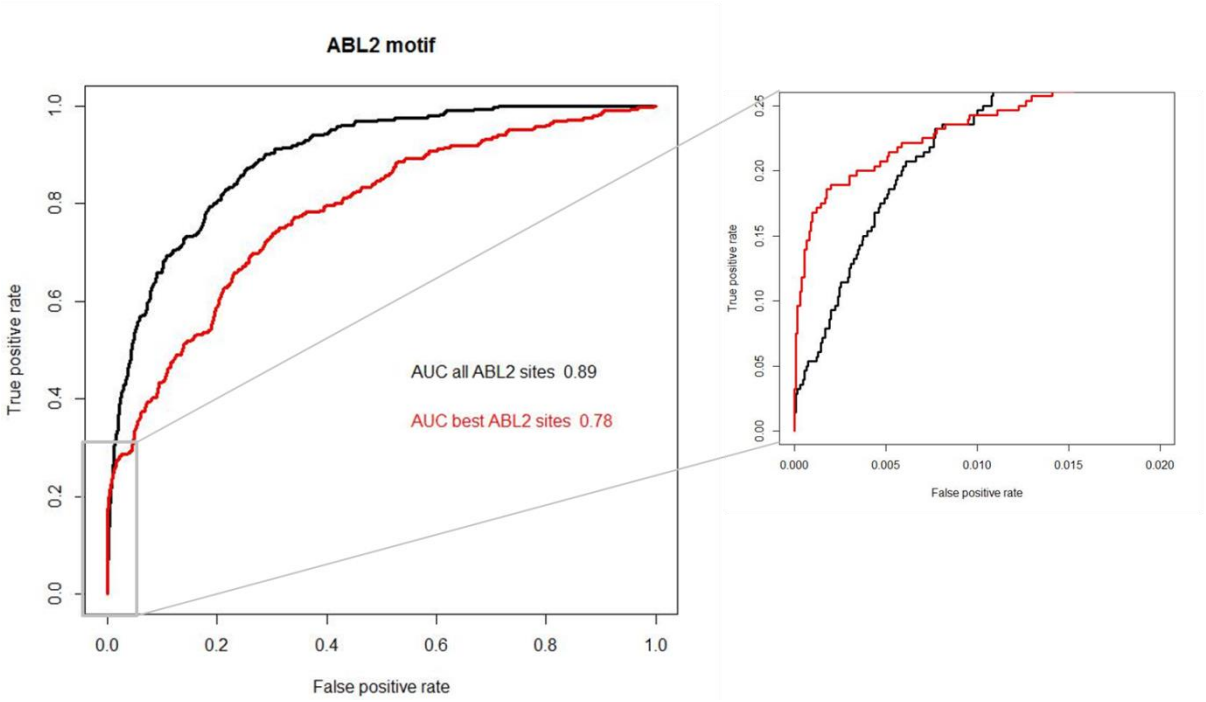


Figure 28: Comparison in performance between the original and the refined motif (left ROC plot). The yeast proteome was scored using the motifs whereas true positives are all yeast pY-sites measured or the best scoring 2% thereof for the refined motif. When zooming in on the start of the resulting curves the better performance of the refined motif in finding the best scoring sites is shown (right ROC plot).

3.11. Sequence motif based scoring of human phosphorylation sites

The refined sequence motifs generated from human kinase targets in yeast were used to assign kinases to previously measured, published human phosphorylation sites. A set of 13240 human pY-sites (set "human_pY") was retrieved from the modification resource PhosphoSitePlus (downloaded 2012) and scored using the PWMs retrieved for each NRTK from the iceLogo program as described above. The distribution of scores over all measured human pY-sites was analyzed for all NRTK motifs using density plots (Appendix 6.5) as shown for ABL2 kinase in Figure 29.

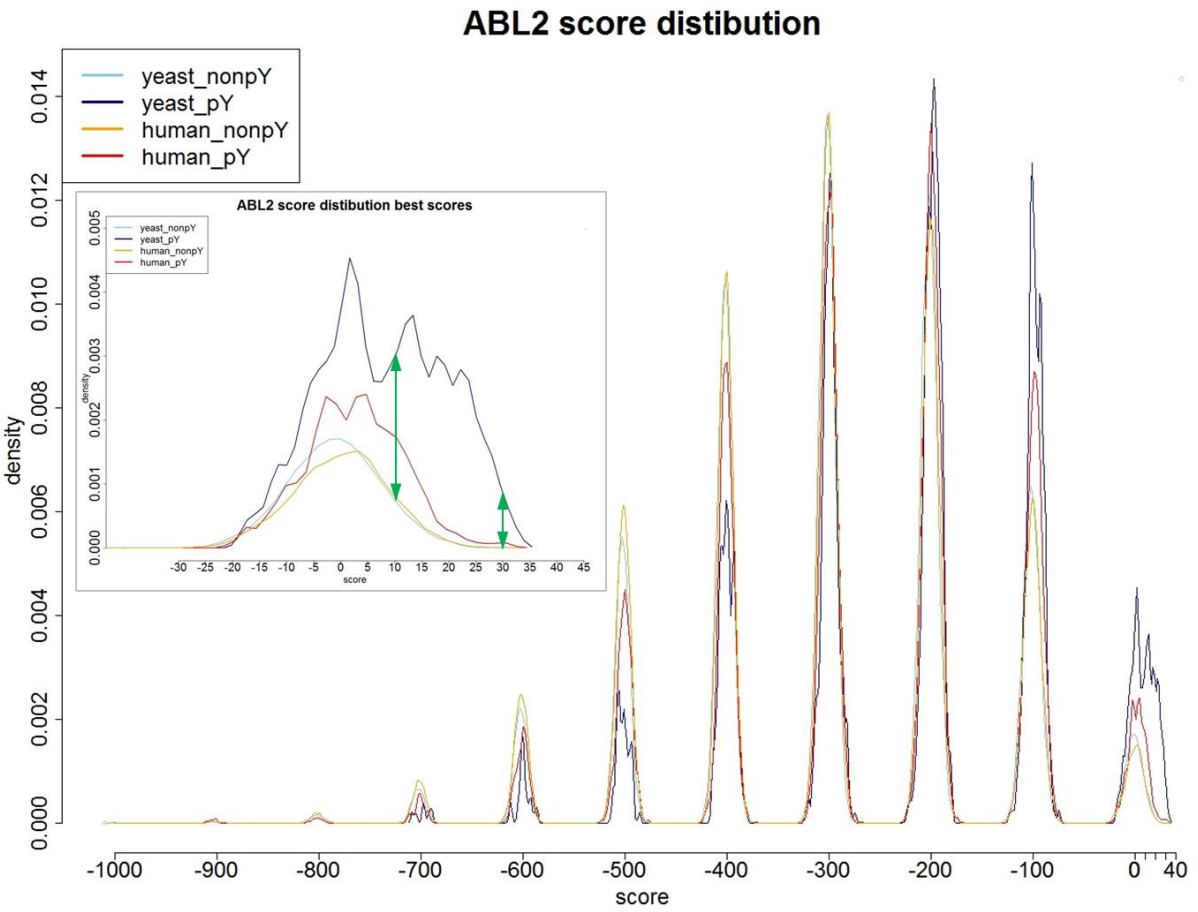


Figure 29: Density plot with score distributions over four 15mer sets (Y flanked by seven amino acids): all 94388 non-modified tyrosine in yeast ("yeast_nonpY", light blue line), all 1433 measured pY-sites in yeast("yeast_pY), dark blue line), all 106230 non-modified tyrosine on human pY-proteins ("human_nonpY", orange line) and all 13240 pY-sites measured in human ("human_pY", red line). The inset highlights the distribution of highest scores ranging from minus 30 to the maximum score.

For comparison, all measured 1433 yeast pY-sites (set "yeast_pY", dark blue line), all 94388 non-modified tyrosine in yeast (set "yeast_nonpY") and all 106230 tyrosine residues on human pY-proteins which were not reported to be phosphorylated (set "human_non_pY") were scored and score distributions plotted. For all NRTK motifs the scores for each set showed a multimodal distribution where the number of sites peaked in windows separated by a score of 100, which was depended on how many strongly under-represented amino acids were present in the 15mer sequence (Appendix 6.5). As presented in Table 5 motif scores ranged on average from -1000 to +28 for all kinases whereas the lowest score was reached for SRMS (-1401) followed by BMX (-1303), FES (-1299) and SYK (-1296) and the highest score for BMX (+35) followed by SRC (32) and ABL2 (31.5). The distribution of the two background sets containing only non-modified sites are very similar (Figure 29, light blue and orange lines). Importantly, there are more yeast and human phosphorylation sites with highest scores compared to yeast and human non-modified sites, respectively (Figure 29 inset, dark blue line vs. light blue line and red line vs. orange line). Predictions based on motif scores are most accurate by choosing a score cut-off where the area-under-the-curve

(AUC) difference between the “yeast_pY” and “yeast_nonpY” sets highest scoring sites is maximal. The maximal AUC difference for sites scored with the ABL2 motif, for instance, lies in between a score range of 10 to 30 (Figure 29 inset, green arrows). In summary, score distribution analysis for all sets suggested that human pY-sites tend to have higher scoring properties than tyrosine residues not reported to be modified and best predictions may be retrieved using a very high score cut-off.

Due to the lack of positive sets of human NRTK phosphorylation sites for motif performance testing it is not possible to assign accuracy values to human motif scored sites using ROC plots. Instead, human sites were scored using the motifs and accuracies assigned which were calculated from the yeast data sets. As the scores are unique for each 15mer there are very few scores in human matching a yeast score exactly. Thus, accuracy values from scored yeast sites were assigned human site scores using the accuracy of the closest, lower yeast motif score as defined by the cross-validation performance testing. After scoring all human phosphorylation sites and assigning accuracy values, the optimal accuracy cut-off as determined previously of 0.995 was chosen across all kinases (Section 3.10). This resulted in the assignment of 1388 human pY-sites in total (approximately 11 percent of all reported human pY-sites) across all 17 kinases. Predicted human kinase target sets were visualized by hierarchical clustering as shown below in Figure 30. All 1388 motif-predicted human kinase substrate relationships are listed in detail in Appendix 6.6.

Between 28 and 74 percent of the motif predictions on human pY-sites are specific for a single kinase (Table 5) observed as non-overlapping clusters as highlighted by red dashed lines for FGR-specific predictions in Figure 30. Moreover, overlapping substrate specificity is predicted as for TNK1 and PTK2, for instance, indicated by shared predictions (Figure 30, green frames). From a kinase view, the majority of predictions comprise a single pY-site per target protein for each NRTK and on average only 3.3 percent of the proteins within each target set contain two or a maximum of three sites predicted for each NRTK. For SRC, SRMS, SYK, and YES1 only a single protein was predicted to be multiply phosphorylated at two sites for the corresponding NRTK (Figure 30, top panel). The kinase with the most proteins predicted to be multiple targeted was PTK2. A total of 13 proteins (almost ten percent of all predicted PTK2 targets) are predicted to be modified on two sites. For both TNK1 and FRK six human proteins are predicted to be phosphorylated on multiple sites by the kinases. Five of these proteins are predicted to be targeted on two tyrosine residues and one protein was predicted for each of the two kinases to be modified on three tyrosine residues. The number of predicted human targets varied between 58 for SRMS and 165 for FER while there were 117 predictions on average. Interestingly, the number of predictions unique to a kinase is between 27 (truncated FYN) and 74 percent (SRMS) and on average 51 percent (Table 5). The lowest amount of unique sites were assigned to the truncated version of FYN, which is deregulated and which was the most active kinase

construct in yeast, resulting in a degenerate FYN motif and hence causing a drop in unique predictions to 27 percent as for normal FYN almost 40 percent unique assignments were retrieved.

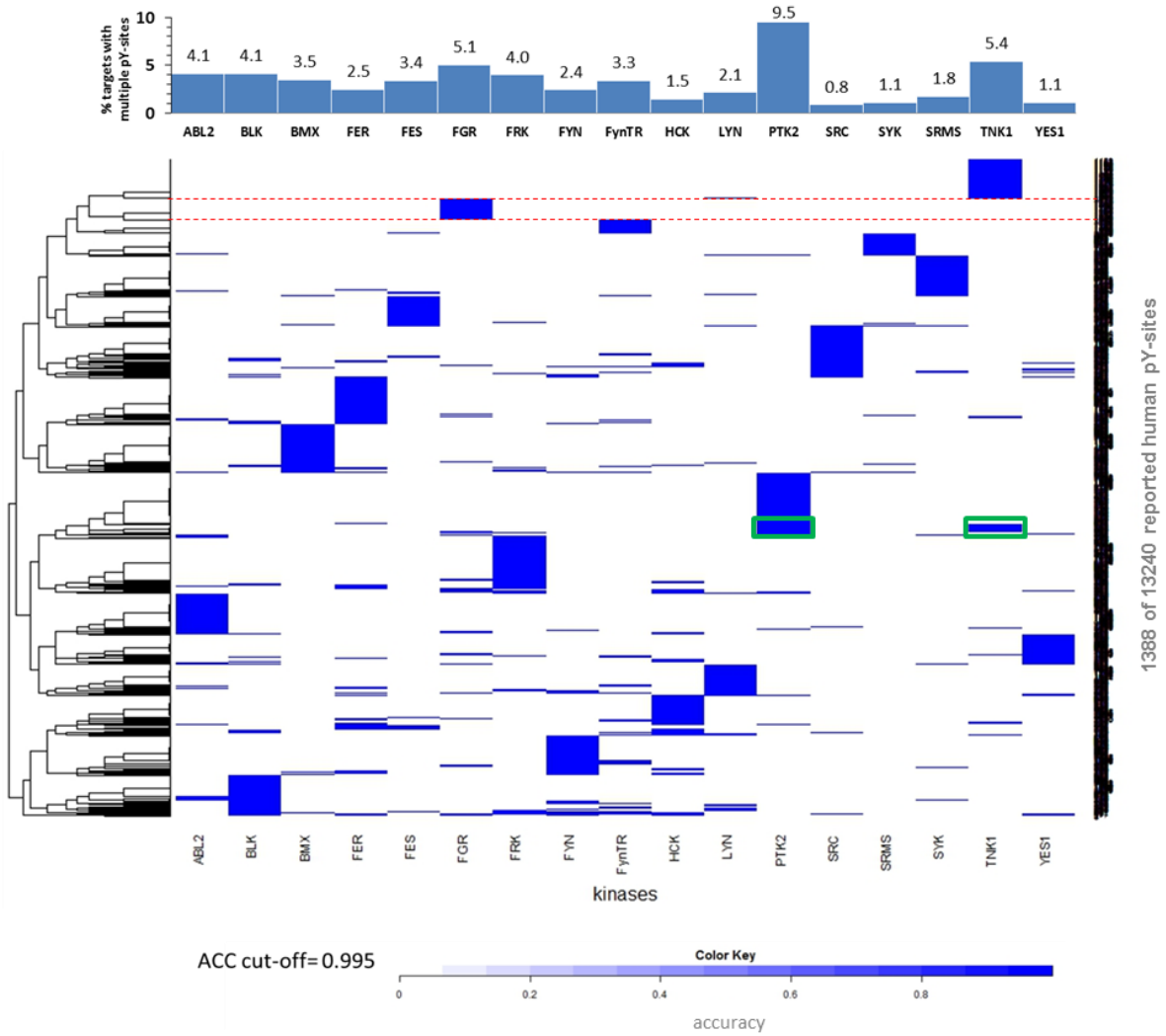


Figure 30: Heatmap showing unique sets of human pY-sites assigned to specific NRTKs by hierarchical clustering using R programming language. Using the accuracy value cut-off 0.995, 1388 of 13,240 reported human pY-sites were matched to NRTKs (Appendix 6.6). The bar graph above represents the percentage of human target proteins with multiple (2 or 3) pY-sites predicted for each kinase.

Overall, roughly half of the 1105 predicted human target proteins were assigned to a single NRTK (Figure 31A) and distribution of kinases per site appears tail-like just as seen for the measurements in yeast outlined in Figure 11. Hence, the distribution suggest a similar NRTK specificity overlap within the human predictions as observed in the yeast measurements. The fact that around half of the targets are uniquely assigned to a kinase showed not only the specificity provided by the motifs, but also the NRTK specificity overlap which makes the assignment of KSRs difficult using human cells. The only human target predicted for nine different kinases (excluding FynTR), on four tyrosine residues, was plakophilin 4 (PKP4) - a signaling molecule important in cell adhesion. Prediction of pY-sites

specific for each NRTK in human is further underlined on a site level by the finding that a single kinase is inferred for approximately 74 percent of all predicted pY-sites (Figure 31B). In this regard, there were only few human proteins predicted to be targeted on a single tyrosine by several NRTKs and only few targets predicted to be modified on several tyrosine mutually exclusive by specific NRTKs. Indeed, only two sites were predicted to be modified by a total of eight NRTKs: Y181 of cyclin-dependent kinase 16 (CDK16) was strikingly predicted to be targeted by all seven tested SFKs and FER, and Y203 of CDK17 by six SFKs and closely related FRK and FER. In contrast, breast cancer anti-estrogen resistance 1 (BCAR1), tensin 1 (TNS1), and neural precursor cell expressed, developmentally down-regulated 9 (NEDD9) were predicted to be targeted on up to nine tyrosine each by mainly one specific NRTK. Excluding FynTR, there were only four predicted targets to be phosphorylated on three motif matching sites by the same kinase: Y128, Y234, and Y362 of BCAR1 by ABL2 and Y179, Y287, and Y327 of BCAR1 by BLK; Y178, Y492, and Y493 of zeta-chain (TCR) associated protein kinase 70kDa (ZAP70) by FRK and Y341, Y347, and Y357 of the exclusively for TNK1 motif predicted target heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1). The latter finding that HNRNPA1 may be targeted exclusively by TNK1 on three motif information carrying tyrosine in close proximity may suggest a pivotal role of TNK1 in HNRNPA1 regulation.

Gene symbol	Predicted human targets	Unique (%)	Reported human targets*	Score range for all human pY-sites
ABL2	129	69 (54)	160	[-910.69 , 31.39]
BLK	128	43 (34)	6	[-999.99 , 28.54]
BMX	120	77 (64)	0	[-1303.60 , 35.23]
FER	165	75 (46)	15	[-1105.22 , 31.32]
FES	92	53 (58)	9	[-1298.98 , 17.43]
FGR	104	45 (43)	12	[-607.58 , 30.19]
FRK	158	84 (53)	3	[-997.37 , 27.64]
FYN	129	51 (40)	161	[-995.12 , 25.43]
FYNΔ	94	26 (28)	NA	[-703.47 , 28.88]
HCK	140	46 (33)	25	[-995.27 , 27.44]
LYN	96	43 (45)	90	[-1005 , 28]
PTK2	150	106 (71)	35	[-1199.71 , 28.80]
SRC	120	55 (46)	543	[-1202.39 , 32.46]
SRMS	58	43 (74)	0	[-1401.41 , -182.54]
SYK	95	70 (74)	73	[-1295.80 , 25.68]
TNK1	119	81 (68)	0	[-707.40 , 25.02]
YES1	94	39 (42)	9	[-904.22 , 28.57]

Table 5: Number of predicted human targets, reported human targets (*from phosphositeplus.org), predicted human targets unique for each NRTK, and the score range over all 13240 human pY-sites for each refined sequence motif. NA = not applicable.

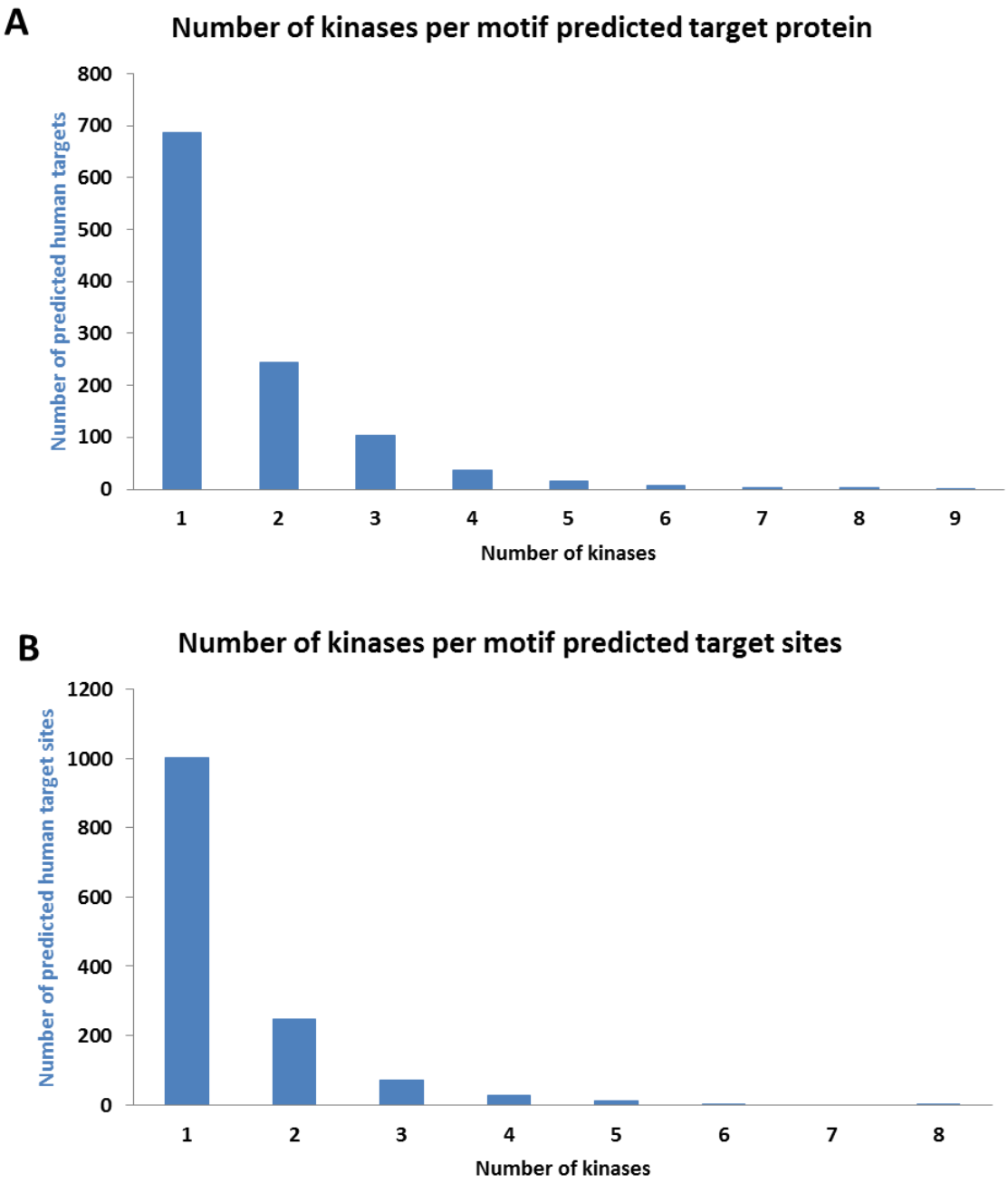


Figure 31: (A) Number kinases versus number of motif predicted human target proteins and (B) number of kinases versus number of motif predicted human phosphorylation sites. The truncated version of FYN (FynTR) was excluded from this analysis.

3.12. Validation of human kinase-substrate relationships based on literature mining

For each kinase all refined motif predicted human phosphorylation sites were queried against all known kinase substrates retrieved from phosphositeplus.org (downloaded 25.02.2015). The number of reported human pY-sites varied greatly from no reported sites for the rather previously little examined NRTKs BMX, SRMS, and TNK1 to a maximum of 543 sites for highly studied SRC (Table 5). The query additionally allowed for validation via orthologous sites reported for other species such as mouse, rat, and chicken. As summarized in Table 6 an overall number of 13 motif predicted human pY-sites were reported previously to be modified by the corresponding kinase. Validated motif predictions by literature will be briefly outlined in the next paragraphs with corresponding literature indicated whereas a more detailed description of selected validated motif predicted human targets is presented in the discussion section (paragraph 4.8.2).

Y221 phosphorylation of v-crk avian sarcoma virus CT10 oncogene homolog (CRK) was previously assigned to ABL2 in human by mutational analysis (Feller et al., 1994). Additionally, site Y239 of the ABL adapter molecule CRK was predicted to be targeted by ABL2 as shown recently *in vitro* by a kinase assay performed by Kumar et al. (2014). Furthermore, both modification site flanking sequences of Y220 and Y232 of mouse Dab, reelin signal transducer, homolog 1 (DAB1) which forms a protein complex with CRK in Reelin-signaling (Ballif et al., 2004) were reported to be targeted by ABL2 (Pramatarova et al., 2003) and have an identical position in their predicted human counterparts. Additionally, phosphorylation of Y36 of mouse Ras and Rab interactor 1 (RIN1), corresponding to predicted Y35 in human RIN1, was assigned to ABL2 previously (Hu et al., 2005). Lastly, prediction of ABL2 targeting of Y171 of human LIM and SH3 protein 1 (LASP-1) is in agreement with previous reports (Lin et al., 2004).

The number of 161 reported sites for FYN was the similar to the number of 160 assigned sites for ABL2 (Table 5). Compared to a total of five pY-sites predicted for ABL2, a total of six pY-sites predicted for FYN were assigned to the kinases previously. Predicted modification of Y487 of B-cell receptor regulator “cluster of differentiation 5” or “CD5 molecule” (CD5) by FYN was shown in an *in vitro* kinase assay testing FYN kinase domain activity on mutated versions of the site containing C-terminal half of CD5 (Vila et al., 2001). Furthermore, predicted phosphorylation of Y15 of mouse cyclin-dependent kinase 5 (CDK5) by FYN was demonstrated using a CDK5 Y15-recognizing antibody probing immuno-precipitate of HEK293T upon co-expression of GST-tagged constructs of mutated and constitutively active mouse Fyn and mouse CDK5 (Sasaki et al., 2002). The pY-site and flanking residues (EKIGEGTYGTVFKAK) are conserved between rat, mouse, and human and moreover; highly

similar to Y15 in homologs CDK1 (EKIGEGTYGVVYKGR), CDK2 (EKIGEGTYGVVYKAR), and CDK3 (EKIGEGTYGVVYKAK) which were also predicted for FYN by a little higher motif score, but without experimental evidence in literature. Motif predicted FYN phosphorylation of Y713 of platelet/endothelial cell adhesion molecule 1 (PECAM1), a protein important in surface adhesion, was determined in vitro Cochrane et al. (2000) by incubating phage displaying PECAM1 fragments with FYN and subsequent characterized to inhibit the interaction with a SH2 binding partner by competition assays using synthesized peptides with and without phosphorylated tyrosine. The tyrosine is further predicted to be modified by BLK, FRK, and LYN. Interestingly, PECAM1 was tyrosine phosphorylated upon aggregation of Fc fragment of IgE, high affinity I, receptor for; gamma polypeptide (FCER1G) as determined by stimulation of rat basophilic leukemia cells by antigen-specific IgE and subsequent immuno-blotting using a PECAM1 specific antibody (Sagawa et al., 1997). FCER1G itself was shown to be modified on two tyrosine residues by LYN which leads to recruitment and activation of SYK in murine in bone marrow-derived mast cells as revealed by in vitro kinase assays and mass spectrometry measurements (Yamashita et al., 2008). LYN targeting of human Y65 of FCER1G was motif predicted whereas the Y65 appears to be more conserved in human FCER1G than the surrounding, especially the amino acids N-terminal to the modification site (Table 6 column 8 versus column 10). Moreover, Y65 of FCER1G matched also the motifs of FER, FRK, SRC, and YES1. Another protein involved in immune cell signaling is FYN binding protein (FYB), also called adhesion and degranulation promoting adapter protein (ADAP), which was reported by Sylvester et al. (2010) to interact with binding partners in a FYN dependent manner as determined using a modified Yeast-2-Hybrid system. *In vitro* phosphorylation of FYB with FYN and subsequent mass spectrometry identified Y571 together with further modified FYB residues including Y559 to be modified by FYN. Tyrosine phosphorylation of both Y557 and Y559 was shown to have an effect on cell adhesion and migration of stimulated Jurkat T cells using comparative expression of phenylalanine substitution FYB mutants. Y571 was the only motif predicted pY-site of FYB for FYN and was additionally predicted for SRC. Furthermore, Y559 was predicted for PTK2 which to my best knowledge has not been reported so far. FYB binds constitutively to the two SH3 domains of SRC Kinase Associated Phosphoprotein 1 (SKAP1) whereas SKAP1 Y295 resides within the binding motif for C-terminal SH3 domain of FYB and phosphorylation of this tyrosine residue by FYN blocks binding as by reported by Duke-Cohan et al. (2006) performing surface plasmon resonance interaction analysis. Y295 was the only residue of SKAP1 predicted to be phosphorylated and the kinase was determined to be FYN.

Furthermore, a YES1 motif predicted kinase-substrate relationship was observed previously. YES1, ABL, SRC, FYN, and LYN targeting of Y311 of mouse protein kinase C, delta (PRKCD) corresponding to Y313 in human PRKCD, was reported by Rybin et al. (2008) via *in vitro* kinase assays and immuno-blotting. Y313 was predicted to match only the YES1 sequence motif and an additional site Y374 was

predicted for FGR which however, has not been observed previously. In regard of SRC motif predictions, sites Y860 of cadherin 2, type 1, N-cadherin (CDH2) and Y758 of coronin 7 (CORO7) were assigned to the kinase previously (Qi et al., 2006, Rybakin et al., 2008). For CDH2, Y860 was predicted for SRC and Y785 for BMX whereas Y860 was identified as the crucial residue in β-catenin dissociation from the N-cadherin adhesion complex during migration of tumor cells (Qi et al., 2006). CORO7 is an ubiquitous WD40-repeat protein that translocates to the Golgi apparatus and participates in the maintenance of the Golgi structure and function Y758 and Y288 was predicted via motif analysis to be modified by SRC and BMX, respectively. Interestingly, Y288 was shown by Rybakin et al. (2006) to be important for CORO7 interaction with an adapter complex required for cargo transport to the Golgi apparatus and prediction of BMX targeting Y288 may suggest an unknown role of BMX in cargo transport in addition to regulation of endocytosis through BMX interaction with an endosomal protein RUFY1 as reported by Yang et al. (2002). Rybakin et al. (2008) also determined in biochemical assays that Y758 phosphorylation by SRC is critical for targeting of CORO7 from the cytosol to the Golgi.

Homology predicted human phosphorylation sites were also queried to be reported in phosphositeplus.org database and not a single orthologous substrate with assigned kinase was identified. However, by manually inspecting literature a recent publication of Fan et al. (2014) was identified who determined ABL kinase in an *in vitro* kinase assay to modify Y301 on pyruvate dehydrogenase alpha 1 (PDHA1). The orthologous site Y321 in PDA1 was targeted solely by ABL2 and hence the resulting prediction based on homology was validated.

Given that the amount of experimentally defined and reported KSRs is very limited, relatively many motif and homology predicted NRTK substrates were validated through previous investigations. For several of the reported kinase-substrate pairs additional NRTKs were predicted to target the modification sites. Due to the fact that many kinase predictions for human pY-sites were reported elsewhere previously increases the confidence that many of the unreported predictions made here are correct.

Kinase	Gene Symbol	Description	Species	UniProt ID	Site	Sequence	Method	Site Hs	Sequence Hs motif
ABL2	CRK	v-crk avian sarcoma virus CT10 oncogene homolog	HUMAN	P46108	Y221	GGPEPGPYAQPSVNT	in vivo, in vitro	Y221	GGPEPGPYAQPSVNT
ABL2	CRK	v-crk avian sarcoma virus CT10 oncogene homolog	MOUSE	Q64010	Y221	GGPEPGPYAQPSVNT	in vitro	Y221	GGPEPGPYAQPSVNT
ABL2	CRK	v-crk avian sarcoma virus CT10 oncogene homolog	CHICKEN	Q04929	Y222	GGPEPGPYAQPSINT	in vivo, in vitro	Y221	GGPEPGPYAQPSVNT
ABL2	CRK	v-crk avian sarcoma virus CT10 oncogene homolog	HUMAN	P46108	Y239*	NLQNGPIYARVIQKR	In vitro	Y239	NLQNGPIYARVIQKR
ABL2	DAB1	Dab, reelin signal transducer, homolog 1	MOUSE	P97318	Y220	PETEENIYQVPTSQK	in vitro	Y220	PETEENIYQVPTSQK
ABL2	DAB1	Dab, reelin signal transducer, homolog 1	MOUSE	P97318	Y232	SQKKEGVYDVPKSQP	in vitro	Y232	SQKKEGVYDVPKSQP
ABL2	LASP1	LIM and SH3 protein 1	HUMAN	Q14847	Y171	IPTSAPVYQQPQQQP	in vivo, in vitro	Y171	IPTSAPVYQQPQQQP
ABL2	PDHA1	pyruvate dehydrogenase α1	HUMAN	P08559	Y301*	MSDPGVSYRTREEIQ	In vitro	Y301	homology predicted
ABL2	RIN1	Ras and Rab interactor 1	MOUSE	Q921Q7	Y35	KPSTDPLYDTPDTRG	in vivo	Y36	KPAQDPLYDVPNASG
FYN	CD5	CD5 molecule	HUMAN	P06127	Y487	DNSSDSDYDLHGAQR	in vitro	Y487	DNSSDSDYDLHGAQR
FYN	CDK5	cyclin-dependent kinase 5	RAT	Q03114	Y15	EKIGEGTYGTVFKAK	in vivo	Y15	EKIGEGTYGTVFKAK
FYN	CDK5	cyclin-dependent kinase 5	MOUSE	P49615	Y15	EKIGEGTYGTVFKAK	in vivo	Y15	EKIGEGTYGTVFKAK
FYN	CDK5	cyclin-dependent kinase 5	MONKEY	F6W6J7	Y15*	EKIGEGTYGTVFKAK	in vivo	Y15	EKIGEGTYGTVFKAK
FYN	PECAM1	platelet/endothelial cell adhesion molecule 1	HUMAN	P16284	Y713	KKDTETVYSEVRKAV	in vitro	Y713	KKDTETVYSEVRKAV
FYN	SKAP1	src kinase associated phosphoprotein 1	HUMAN	Q86WV1	Y295	TRRKGVDYASYYQGL	in vivo	Y295	TRRKGVDYASYYQGL
FYN	FYB	FYN binding protein	HUMAN	O15117	Y571	TTAVEIDYDSLKLKK	in vitro	Y571	TTAVEIDYDSLKLKK
LYN	FCER1G	Fc fragment of IgE, receptor for; gamma polypeptide	MOUSE	P20491	Y65	REKADAVYTGLNTRS	in vitro	Y65	YEKSDGVYTGLSTRN
SRC	CDH2	cadherin 2, type 1, N-cadherin (neuronal)	HUMAN	P19022	Y860	DSLLVFDYEGSGSTA	in vitro	Y860	DSLLVFDYEGSGSTA
SRC	CORO7	coronin 7	HUMAN	P57737	Y758	GDTRVFLYELLPESP	in vitro	Y758	GDTRVFLYELLPESP
SRC	CDK5	cyclin-dependent kinase 5	MONKEY	F6W6J7	Y15*	EKIGEGTYGTVFKAK	in vivo	Y15	EKIGEGTYGTVFKAK
YES1	PRKCD	protein kinase C, delta	MOUSE	P28867	Y311	TTESVGIYQGFEKKP	in vitro	Y313	SSEPVGIYQGFEKKT

Table 6: Motif and homology predicted human phosphorylation sites reported to be targeted by the corresponding kinase in phosphositeplus.org database and elsewhere. From left to right: Reported kinase-substrate pair, substrate description, reported organism, ID, reported pY-site and sequence, predicted human pY-site (PS hs) and sequence. Asterisk indicate that the kinase was not assigned in phosphositeplus.org, but identified in literature elsewhere. Corresponding literature for all kinase-site pairs is presented in text.

3.13. Kinase-substrate relationships predicted by both homology and linear sequence motifs

Reported human phosphorylation sites predicted to be targeted via both homology to yeast targets and linear sequence motifs were retrieved. A total of 13 human pY-sites were predicted by both prediction methods to be targeted by some NRTKs (Table 7). For nine of those sites at least one kinase-substrate pair was predicted by both methods. For instance, Y15 of heat shock 70 kDa protein 8 (HSPA8) was homology predicted to be targeted by FGR and FYN and sequence motif predicted to be targeted by FYN and YES1. The fact that FYN was targeting the homologous tyrosine in yeast and additionally, the pY-site matches the FYN sequence motif which may provide additional evidence for correct kinase assignment. In total, 19 kinase-substrate pairs were predicted by both methods (marked in red in Table 7).

3.14. Experimental validation of kinase-substrate pair predictions in human

3.14.1. Selection of kinase-substrate pairs

Predictions based on bioinformatics analysis were exemplary validated experimentally. Potential human kinase substrates were expressed in yeast cells together with the corresponding NRTK. Substrates were immuno-purified and phosphorylation was assessed via mass spectrometry. Kinases were chosen according to domain family and motif characteristics. ABL2 was chosen as a proline directed, well-studied kinase and FGR as the most active SFK in yeast. Predicted human targets were selected based on homology and motif scoring. Based on homology of 63 pY-sites conserved in human in total, 13 pY-sites were predicted for ABL2 and 21 pY-sites for FGR to be targeted. For sequence motif predicted target set, the 20 best scoring sites were selected for each kinase. A prerequisite for selection of both homology and motif predicted sites was that they were reported in at least two publications or measurements (as reported on phosphositeplus.org) and whether the site is present within the translated sequence of the available gateway cDNA clones. The clone availability led to a selection of 26 clones for ABL2 (15 motif- and 11 homology-predicted) and 39 clones for FGR (12 motif- and 27 homology-predicted).

Entrez GeneID	Symbol	Description	Site	Homology predicted kinases	Motif predicted kinases
71	ACTG1	Actin, gamma 1	Y53	ABL2, BMX, PTK2, TNK1	PTK2
71	ACTG1	Actin, gamma 1	Y166	ABL2, BLK, FGR, FYN, HCK, SRC, YES1	BLK, FER, FYN, SRC, YES1
1655	DDX5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5	Y202	FYN, LYN	FER, FGR, FYN, LYN
55140	ELP3	Elongation protein 3 homolog (yeast)	Y329	BLK, FER, FGR, HCK, TNK1, YES1	BLK, FER, FGR, FYN, HCK, YES1
2027	ENO3	Enolase 3 (beta, muscle)	Y131	FER, FES, FYN, HCK	ABL2, LYN
3312	HSPA8	Heat shock 70kDa protein 8	Y15	FGR, FYN	FYN, YES1
51602	NOP5/NOP58	NOP58 ribonucleoprotein homolog (yeast)	Y342	ABL2, BLK, FGR, LYN	ABL2
5688	PSMA7	Proteasome (prosome, macropain) subunit, alpha type, 7	Y153	FER, FGR, FYN, HCK, YES1	FGR, PTK2, YES1
6157	RPL27A	Ribosomal protein L27a	Y52	BLK, TNK1	SYK
6132	RPL8	Ribosomal protein L8; ribosomal protein L8 pseudogene 2	Y133	BMX, FES, FRK, TNK1	FES, TNK1
6449	SGTA	Small glutamine-rich tetratricopeptide repeat (TPR)-containing, alpha	Y158	HCK, TNK1	FYN
7086	TKT	Transketolase	Y275	HCK	ABL2, BLK
56897	WRNIP1	Werner helicase interacting protein 1	Y534	BMX	BMX, FRK, HCK

Table 7: Human pY-sites predicted to be targeted via homology and sequence motifs. Kinases predicted by both methods are marked in red.

3.14.2. Experimental workflow

Based on a yeast chromatin immuno-precipitation protocol (Grably and Engelberg, 2010) an experimental workflow was developed for human NRTK target enrichment from yeast lysate via IgG conjugated agarose beads (Figure 32).

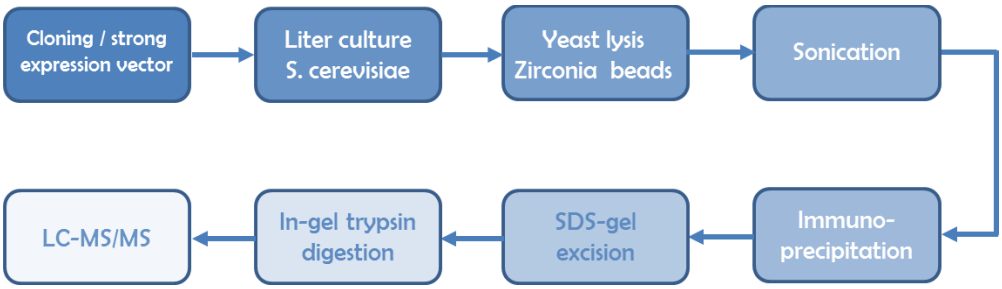


Figure 32: Workflow enrichment of human predicted targets from yeast lysate. A C-terminal *S. aureus* Protein A fusion enables binding to IgG conjugated agarose beads. Explanation in text.

In brief, clones were shuttled in to a gateway yeast expression vector (under control of a GPD promoter) yielding a C-terminally *S. aureus* Protein A tagged fusion protein (Figure 32, step 1). Protein A binds strongly to immuno-globulin and hence can be detected with various antibodies and can be used for immune-precipitation. Thus, expression of the human cDNA in yeast was ensured by analysis of yeast lysate by Western blotting and HPR-linked rabbit IgG antibody. The protein levels in the lysates varied greatly - from proteins showing a very strong band on an western blot to proteins that could not be detected with this method. cDNA clones of putative kinase substrates were selected according to the strongest Western blotting signal. After expression of selected clones together with the predicted kinase in yeast liter cultures cells were harvested by centrifugation, all growth medium removed, and vials containing approximately 1 ml dry pellet frozen to -80°C (Figure 32, step 2). Five vials or more, depending on protein expression levels, were lysed using HEPES lysis buffer and Zirconia beads. Pellets were additionally sonicated in order to open the nuclei to ensure the protein of interest is extracted (step 3 +4). After over-night immuno-precipitation using rabbit IgG conjugated agarose beads and washing steps, proteins were eluted from the beads by addition of SDS-gel loading buffer and heating to 95°C (Figure 32, step 5). The sample was applied to an SDS-gel and, if visible, coomassie blue stained bands were cut out. The gel pieces were grinded using a pestle in order to improve subsequent in-gel trypsin digestion by increasing the protein accessibility for the enzyme (Figure 32, step 6 + 7). After further preparation steps, enriched peptides were applied to a Q-Exactive (Thermo Scientific Inc.) mass spectrometer with optimized settings shown below (Table 8)

using target preparation specific peptide sequence inclusion list (Figure 32, step 8). The sequences in the inclusion list are the first to be searched for in each mass scan performed. The MaxQuant mapping software (Cox and Mann, 2008) outputs a mass spectrum and an evidence file with peptide sequences, mapping identifiers, modifications, probability scores, intensities, retention times, mass errors, number of isotopic peaks, and few other parameters.

Settings	**MS**	**MS/MS**
Collision energy (CE)	27 eV	
Run time	210 minutes	
Resolution	70000	70000
Ion accumulation	120 ms	120 ms
Maximum no. of ions	1 mio	0.5 mio

Table 8: Mass spectrometry settings Q-Exactive

Successful human target enrichments from yeast lysate are shown in Figure 33. The homology predicted ABL2 target phosphoglycerate mutase 1 (PGAM1), the homology predicted FGR target phosphoglycerate kinase 1 (PGK1) and motif predicted FGR target eukaryotic initiation factor 2 subunit 1 (EIF2S1) were strongly expressed in yeast and could be sufficiently enriched with the procedure presented above. Expression was detected as strong (double) bands with the expected size of the fusion proteins by Western blotting and Coomassie Blue staining. The target protein bands were clearly diminished in the unbound fraction and not detectable in the first wash. After applying five percent of the beads to the SDS-gel a strong enrichment of the target on the IgG-beads was detected. Further bands detected below on the Western blots are due to the immuno-reactivity of the unspecific rabbit anti-goat antibody as it may also bind to other co-precipitated proteins.

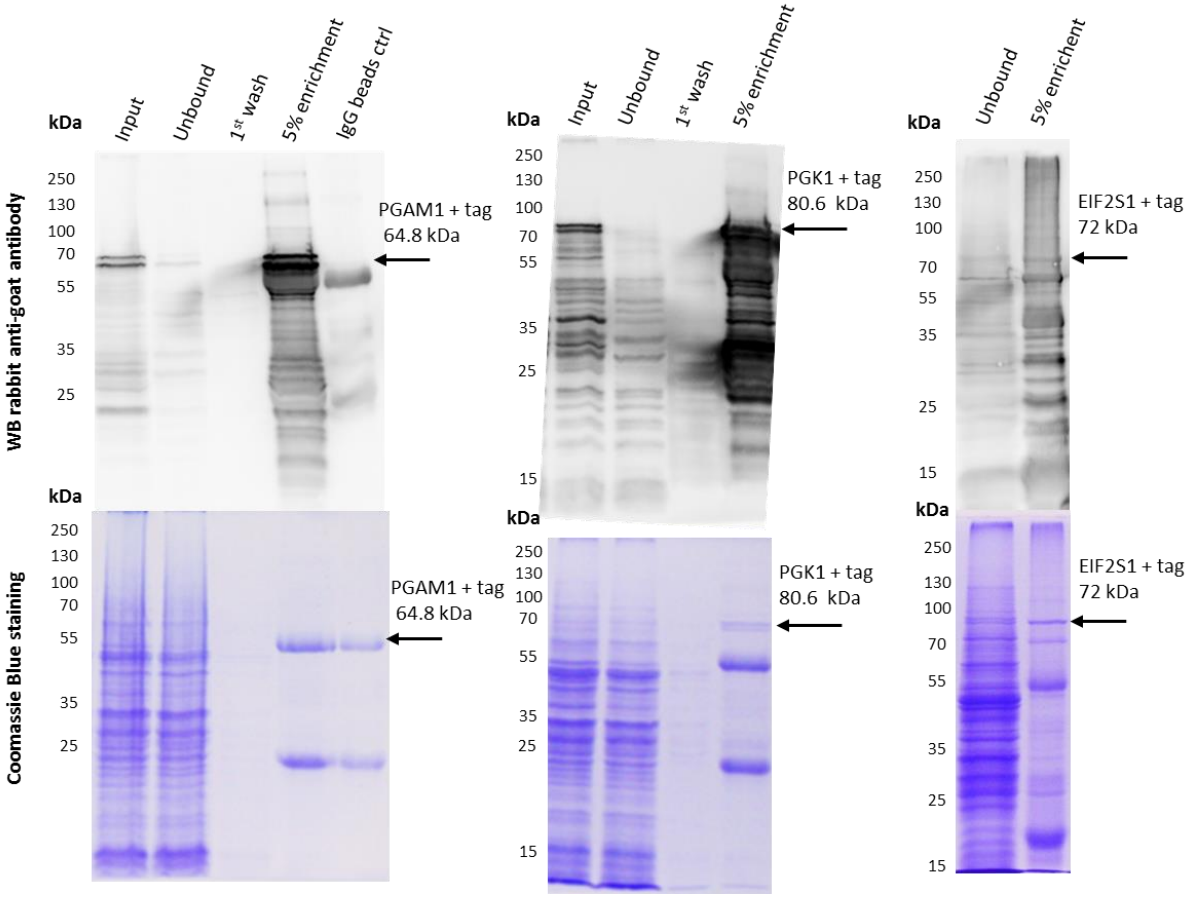


Figure 33: Enrichments for homology predicted ABL2 target PGAM1, homology predicted FGR target PGK1, and motif predicted FGR target EIF2S1. Western blotting using rabbit anti-goat HPR-conjugated antibody (above) and corresponding Coomassie Blue staining (below). Arrows indicate the expected protein sizes.

3.14.3. Kinase-substrate pair validations via mass spectrometry

The success of phosphorylation site detection was strongly dependent on human target expression strength in yeast, the efficiency of the protein to be enriched by the developed protocol, and the activity of the co-expressed NRTK. For three predicted proteins showing sufficient expression in yeast phosphorylation sites could be measured and hence the kinase confidently assigned. Only for those three proteins a Coomassie Blue stained band could be detected indicating that micro-gram of protein was required for successful identification of the modification site via mass spectrometry (Figure 33). The band for PGAM1 at approximately 65 kDa was hardly observable at a little higher weight than the heavy IgG chain however, clearly visible for PGK1 and EIF2S1 (Figure 33). One site, Y92 on PGAM1, was predicted by homology to be phosphorylated by ABL2 and another two sites, Y76 and Y196, on PGK1 were predicted by homology for FGR (Table 9). Another successfully purified protein harbored Y150 on EIF2S1 which was predicted via motif for FGR (Table 9).

Symbol	Description	Entrez GeneID	Peptide + miscleavage	Site	Kinase	Prediction	WB signal	#ref
PGAM1	phosphoglycerate mutase 1	5223	H**Y**GGLTGLNK + AETAAK	Y92	ABL2	homology	strong	1937
EIF2S1	Eukaryotic translation initiation factor 2 subunit 1	1965	YK + RPGYGA**Y**DAFK	Y150	FGR	motif	strong	79
EIF2S1	Eukaryotic translation initiation factor 2 subunit 1	1965	YK + RPG**Y**GAYDAFK	Y147	FGR	-	strong	12
PGK1	phosphoglycerate kinase 1	5230	**Y**SLEPVAVELK	Y76	FGR	homology	strong	53
PGK1	phosphoglycerate kinase 1	5230	ELN**Y**FAK	Y196	FGR	homology	strong	406

Table 9: Description of predicted and observed human kinase targets. #ref stands for number of previously reported measurements. The peptide containing Y196 of PDK1 was only observed in non-phosphorylated form.

For each of the peptides containing putative phosphorylation sites of PGAM1, PGK1, and EIF21S1 a consensus mass spectrum was retrieved from peptideatlas.org and compared to the measured mass spectra. The spectra contain b-ion and y-ion peaks representing fragmentation patterns of the respective precursor peptide enabling sequence determination from both peptide ends. Figure 34 shows the measured mass spectrum for the homology predicted site Y92 on PGAM1 tested with ABL2 together with the consensus mass spectrum for the phosphorylated peptide from peptideatlas.org. In the consensus spectrum all b-ions (b5 to b9, blue peaks) show the higher mass of corresponding fragment ions due to phosphorylation of the tyrosine. The addition of a phosphate to the tyrosine on the y9 fragment also causes a mass shift of approximately 80 Da. All b-ions, apart b6, and the y9-ion with higher masses due to phosphorylation reported in the consensus spectrum were detected. Additionally, the y8-ion was identified having the expected mass without phosphorylation.

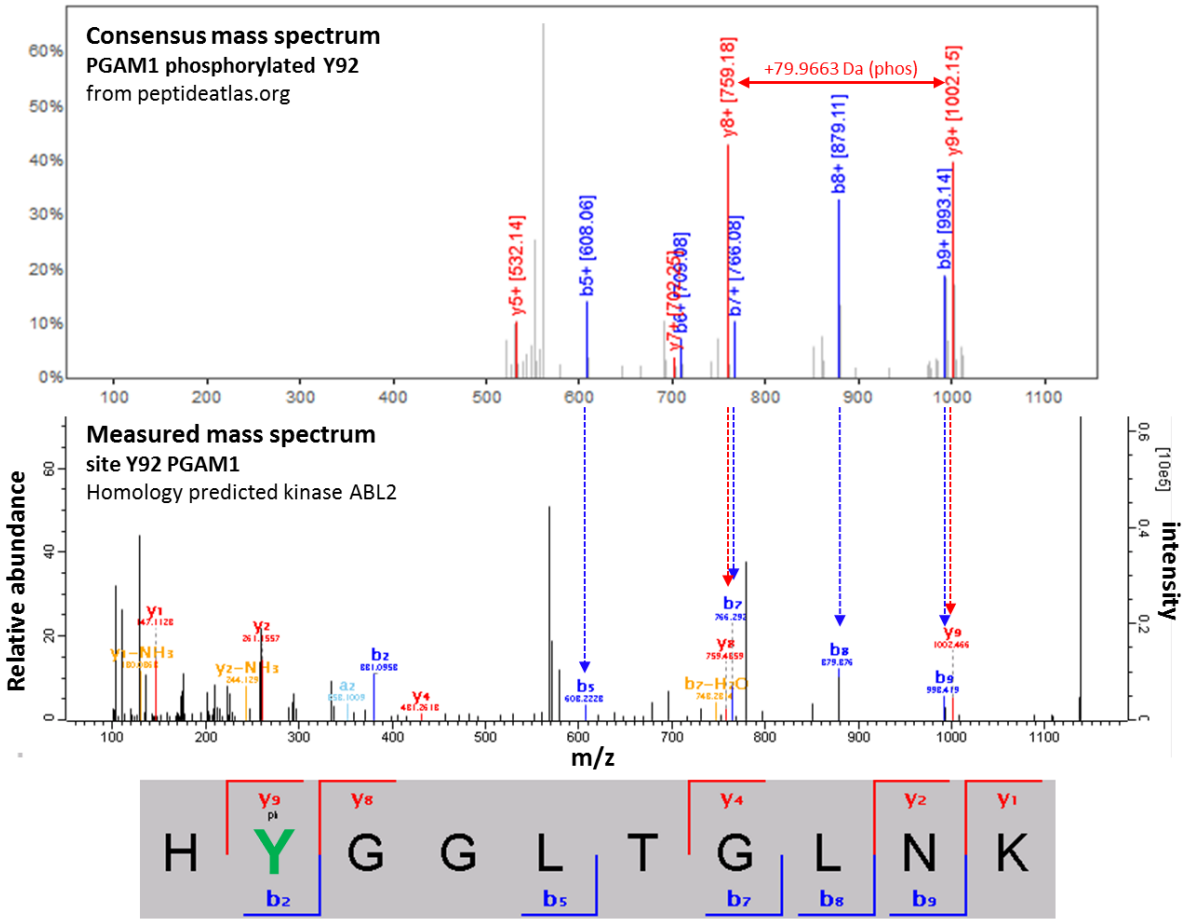


Figure 34: Consensus and measured mass spectrum for site Y92 on PGAM1 (predicted kinase ABL2) and peptide amino acid sequence with detected y (red) and b (blue) fragmentation ions (below). Identified peaks identical to the consensus phosphorylated peptide mass spectrum from peptideatlas.org are indicated by dashed arrows. All b-ions and the y9 ion are covering the modification site.

For the PGK1 two phosphorylation sites were homology predicted for FGR (Table 9). The protein was highly expressed, a strong Coomassie Blue stained gel band could be excised, and many spectra were

observed. Only one consensus motifs without phosphorylation could be retrieved for both peptides harboring the sites (Figure 35). For the Y196 containing peptide ELNYFAK the consensus motif additionally includes a water loss at the N-terminal glutamate which was considered for direct spectra comparisons. The complete series of y-ions were measured covering the entire Y76 containing peptide YSLEPVAVELK (Figure 35A) and are identical in mass to the y-ions in the consensus spectrum. The phosphorylation site was localized by the distance of the ion y10 to the mass of the precursor ion, i.e. by the overall mass of the modified peptide. The fragment y10 ion excluded the possibility that the tyrosine flanking serine was phosphorylated leading to the higher total mass of the precursor ion. The other peptide ELNYFAK was detected in vast numbers however, never observed to be phosphorylated. Five y-ions and two b-ions (with accounted water loss) are identical to the peaks in the consensus spectrum and the additional y6 ion suggest correct identification. The captured phosphorylation events indicate that it is possible to predict KSRs based on sequence similarity between yeast and human.

The measured mass spectra presented in Figure 36 represent the phosphorylation of the sequence motif predicted site Y150 on EIF2S1 and the unpredicted modification of a neighboring tyrosine Y147. For the modified peptide RPGYGAYDAFK only a consensus spectrum with N- and C-terminal iTRAQ (isobaric Tags for Relative and Absolute Quantitation) labels was available. By subtracting the label mass from the consensus fragmentation ions enables comparison with the measured peaks. Thus, Ions y1 to y4 are corresponding to consensus peaks and consensus y10, b8 and b10 fragments were measured when subtracting the mass of the label and adding the phosphorylation mass. Hence, in Figure 36 both spectra A and B were identified correctly in comparison to the consensus spectrum with three peaks indicating the phosphorylation. In first the spectrum for Y150 (Figure 36A), a mass shift of 80 kDa could be measured between y4 and y5 and between b6 and b7 and as mentioned further y and b ions with mass shifts strengthen the evidence. In the second spectrum for Y147 (Figure 36B) , a mass shift of 80 kDa was observed between y7 and doubly charged y8 ion whereas as for site Y150 further y and b ions including the mass of the modification were identified. The successful measurements suggest that using the refined sequence motif generated from yeast substrate sequences can be used to predict the kinase for a subset of human phosphorylation sites. The data from the experimental validation are summarized in Table 9.

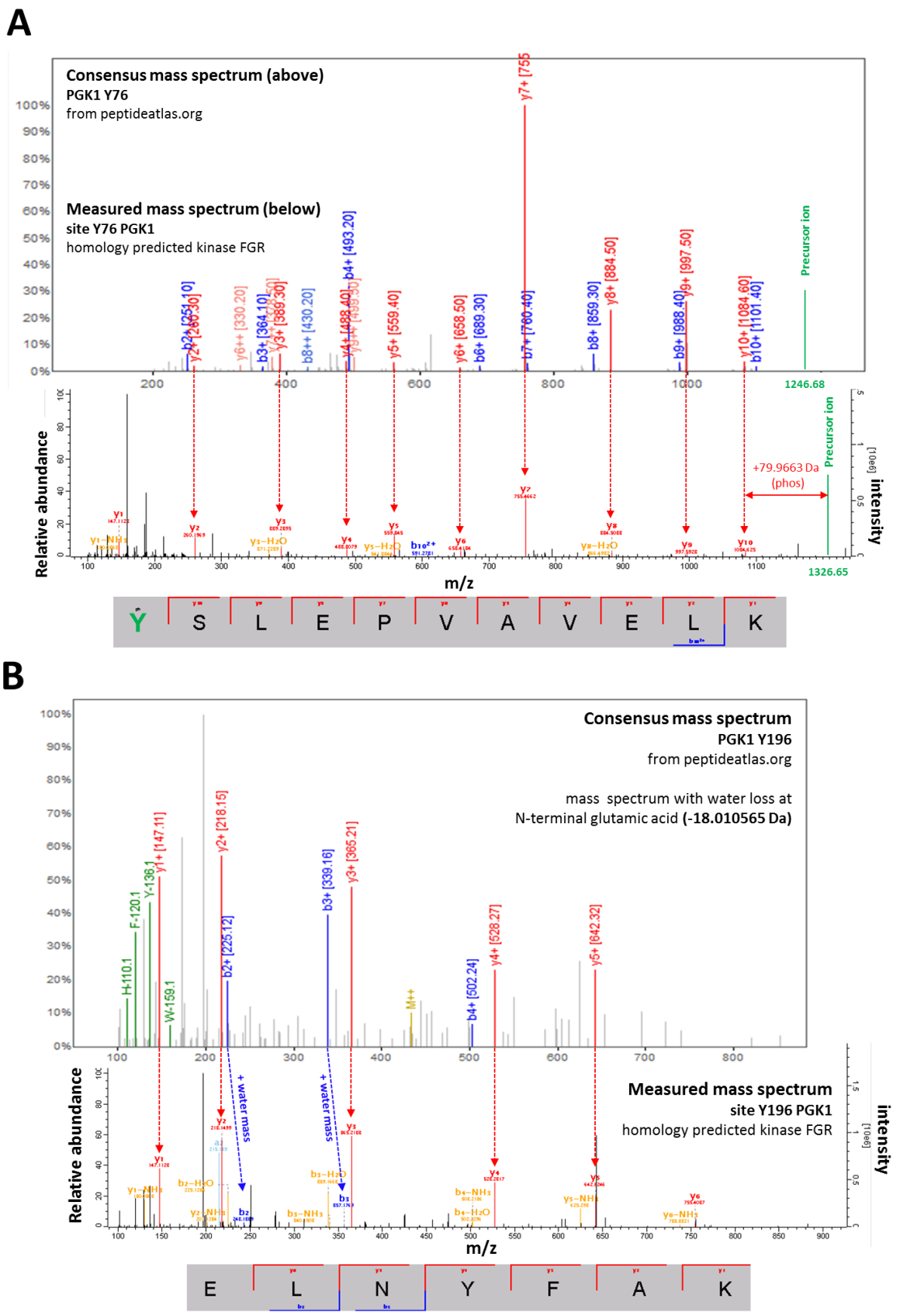


Figure 35: **Measured mass spectrum for sites Y76 and Y196 on PGK1 (predicted kinase FGR) and peptide sequence with detected y (red) and b (blue) fragmentation ions (below). Dashed arrows indicate peaks identical to the consensus phosphorylated peptide mass spectrum from peptideatlas.org are. Y76 phosphorylation is detected by the mass shift from y10 to the precursor mass (green). Y196 is not phosphorylated.**

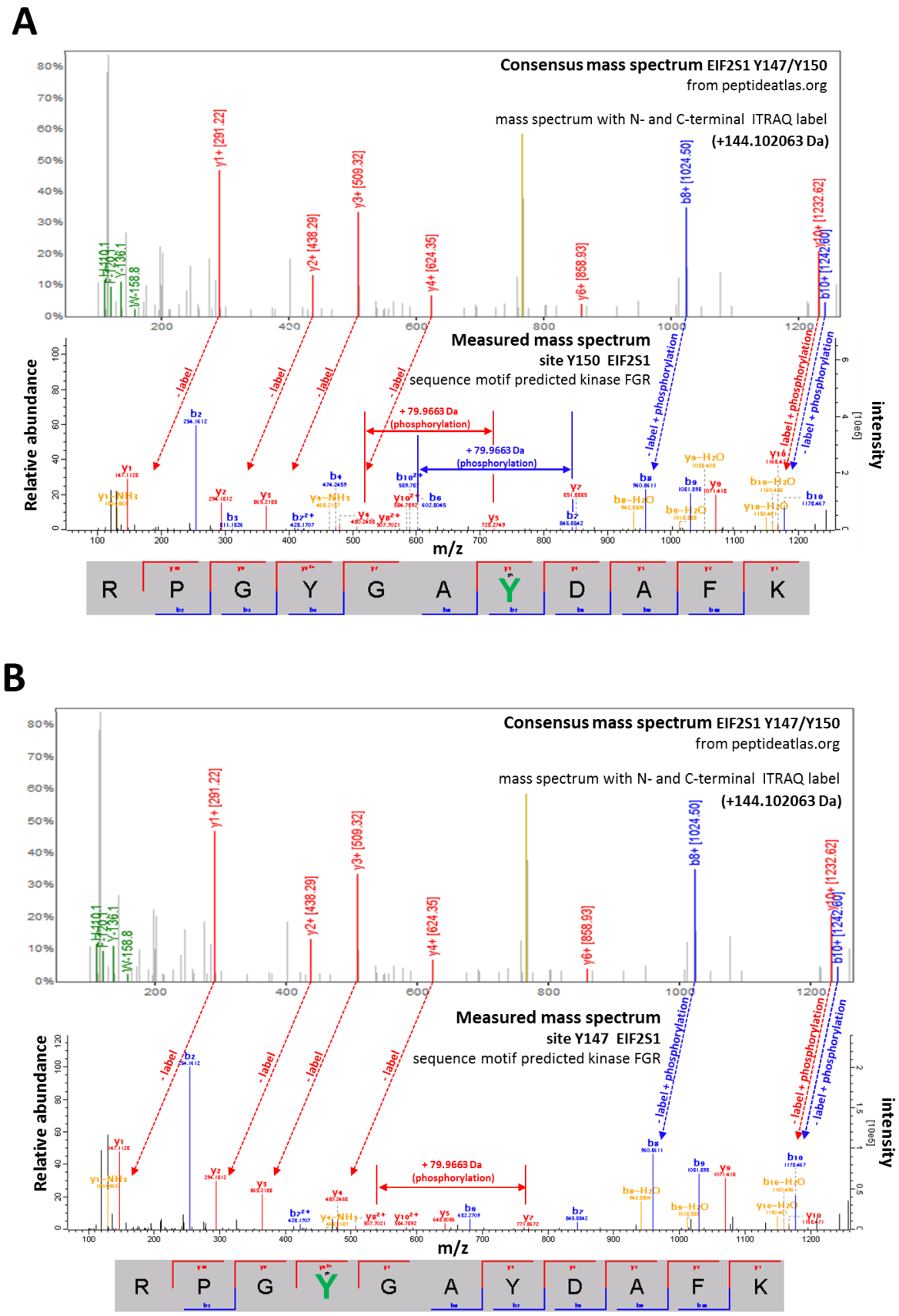


Figure 36: Measured mass spectrum for sites Y150 (above) and Y147 (predicted kinase FGR) and peptide sequence with detected y (red) and b (blue) fragmentation ions (below). Dashed arrows indicate peaks identical to the consensus phosphorylated peptide mass spectrum from peptideatlas.org. All b-ions, apart from b2, and the doubly charged y8, y9 and y10 ions are covering the modification site.

3.15. Kinase targeting in the context of yeast binary PPI networks

Expression of active human kinases in a highly crowded environment enables not only determination of substrate properties such as linear sequence motifs, but also the influence of the cellular environment on substrate targeting on a systems level. Analysis of Protein-Protein Interaction (PPI) network topology of measured kinase targeted yeast proteins in comparison to appropriate null model distributions may reveal regulation on a system-wide level. For network analysis two well defined binary yeast PPI datasets were used. On the one hand, a physical PPI network connecting 5878 nodes with 70529 edges was retrieved from Consensus Path DB version 26 (Kamburov et al., 2013). On the other hand, a probabilistic PPI network was retrieved from STING database at a confidence score cut-off of 0.9 (Snel et al., 2000) connecting 4618 nodes with 45953 edges. In general, both network resources were constructed by integration of different interaction sources. CPDB integrates binary and complex protein-protein, genetic, metabolic, signaling, gene regulatory and drug-target interactions, and biochemical pathways. For the analysis here the CPDB yeast network was filtered and solely binary, physical protein-protein interaction data were retrieved. STRING integrates curated pathway databases, physical protein-protein interaction screens, transcript expression analysis, genomic context, and co-occurrence of interaction partners in publication abstracts. Thus, STRING can capture direct and indirect interactions which could be otherwise missed such as the predicted interactions based on protein scaffolding (Linding et al., 2007). From the probabilistic STRING database, a binary yeast PPI network was obtained using a recommended, highly stringent cut-off with STRING scores greater than 0.9 (a value of zero denotes lowest and a value of 1 highest probability for each interaction). The two networks have very different properties (see below) however, the rational in using two different networks in the analyses was to see how robust results are with respect to the underlying PPI network data.

	target proteins	CPDB yeast nodes	% targets in network	CPDB yeast edges	STRING yeast nodes	% targets in network	STRING yeast edges
Network		5878		70529	4618		45953
ABL2	240	225	93.8	214	225	93.8	195
BLK	225	210	93.3	283	217	96.4	215
BMX	101	91	90.1	66	100	99.0	110
FER	218	212	97.2	195	197	90.4	131
FES	82	74	90.2	53	81	98.8	64
FGR	332	312	94.0	563	313	94.3	546
FRK	202	189	93.6	172	191	94.6	152
FYN	201	184	91.5	266	195	97.0	207
FynTR	341	318	93.3	453	322	94.4	355
HCK	184	171	92.9	163	173	94.0	116
LYN	138	122	88.4	144	134	97.1	238
PTK2	90	83	92.2	66	84	93.3	42
SRC	98	93	94.9	56	93	94.9	27
SRMS	23	21	91.3	7	23	100.0	24
SYK	79	73	92.4	58	77	97.5	69
TNK1	266	248	93.2	304	248	93.2	330
YES1	210	198	94.3	226	201	95.7	124

Table 10: Comparison of target set size to number of targets in the CPDB and STRING networks. Also the number of edges for each NRTK in both networks is presented.

Using Network X (release 1.7), a published network analysis module in PYTHON programming language (Hagberg et al., 2008), different distance measures between kinase targets in the context of two PPI networks were calculated. Initially, the average shortest path (ASP), i.e. the average minimum number of nodes passed by on the shortest direct path between each pair of nodes within the kinase targets sets, was compared to the ASP in randomized networks (Albert, 2005, Watts and Strogatz, 1998). The null hypothesis here was that, when replacing targeted nodes by randomly picked nodes from the entire network and the ASP does not change when calculated between the subset of targets, it is not a characterizing property of the original sub-network. The original network structure was preserved in the randomized networks as only nodes were shuffled maintaining the number of interaction partners, i.e. the same degree. As a measure of significance a Z-score was calculated as a subtraction of the mean of ASPs of 100 randomized networks from the ASP the target set divided by the standard deviation of ASP normal distribution as shown in the following equation:

$$z = \frac{X - \mu}{\sigma};\ \ z = \frac{ASP\_target\_set - mean\ ASP\_randomizations}{standard\ deviation\ ASP\_randomizations}$$

As it can be seen in Figure 37, for each kinase target set the proteins are significantly closer within the interactomes tested than expected by random chance suggesting that processes or protein

complexes are targeted rather than remotely distributed proteins within in the PPI networks. The deregulated FYN construct lacking the SH2 and SH3 domains appears to have targeted proteins which are very close in the network, suggesting that domain interactions may lead to substrate specification and prevention of arbitrary targeting within a local interaction space.

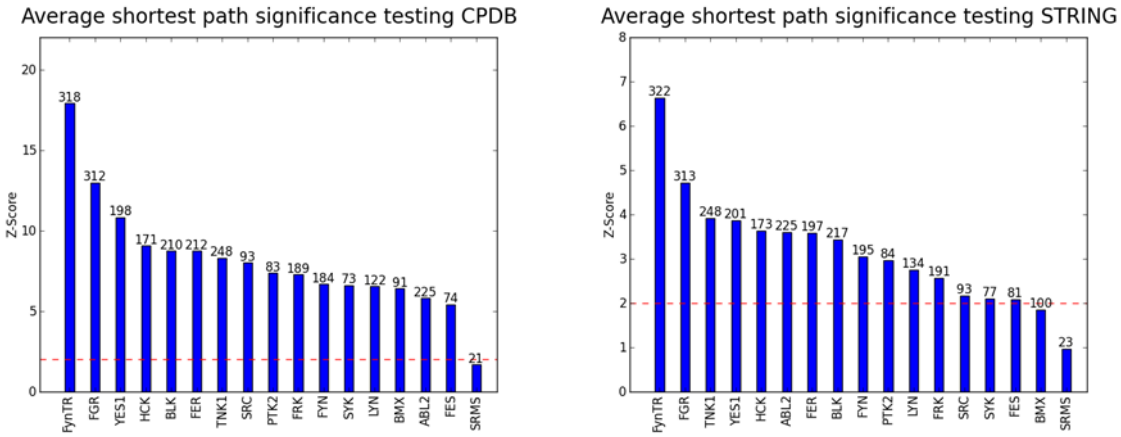


Figure 37: Z-scores for each kinase target set retrieved from node randomization average shortest path significance testing. Bar graphs above the red dashed line at a Z-score of 2 indicate significantly shorter network distance within the kinase target set.

The significance of shorter path distance is approximately three fold more significant in the solely physical PPI network CPDB than in the highly wired STRING network using a high cut-off. Another network property describing node inter-linked is the clustering-coefficient (CC). A CC of 1 means that all neighbors of a node are inter-linked via all possible edges while a CC of 0 means the node neighbors are isolated from its other neighbors (Watts and Strogatz, 1998). The histograms showing network-wide frequencies of clustering coefficients (CC) (Figure 38A) depict the major differences between the two networks. Most nodes in the CPDB network show relatively poor linkage between nodes as CCs are overall small and tail-out towards only a small set of completely inter-linked proteins or complexes (Figure 38 A, left). Oppositely, around 400 nodes were completely interlinked in the binary STRING network retrieved with a very high probability cutoff, depicting most likely protein complexes, and a broad distribution of CC over all nodes was determined (Figure 38 A, right). As a consequence, it was observed that, similar to global CC distributions, kinase target sets within the STRING network have a much higher average median CC than within the CPDB network (Figure 38B). Apart from three kinase target sets, within the STRING network kinase target sets contain minimum and maximum CCs (whiskers) and the first quartile (the value between minimum and median data CC values) was above 0.2 for all target sets. Within the CPDB network only few outliers show CCs above 0.5 and the median CCs are around 0.1 to 0.2. Thus, kinase target sets are more interlinked within the STRING network as compared to the CPDB network. In order to investigate whether the kinase target inter-linkage is significantly higher than expected by random chance, i.e.

whether degree preserved randomly picked nodes from the same network show a similar distribution of CCs, 100 times node-shuffling and Z-transformation was performed as previously for the analysis of average shortest paths.

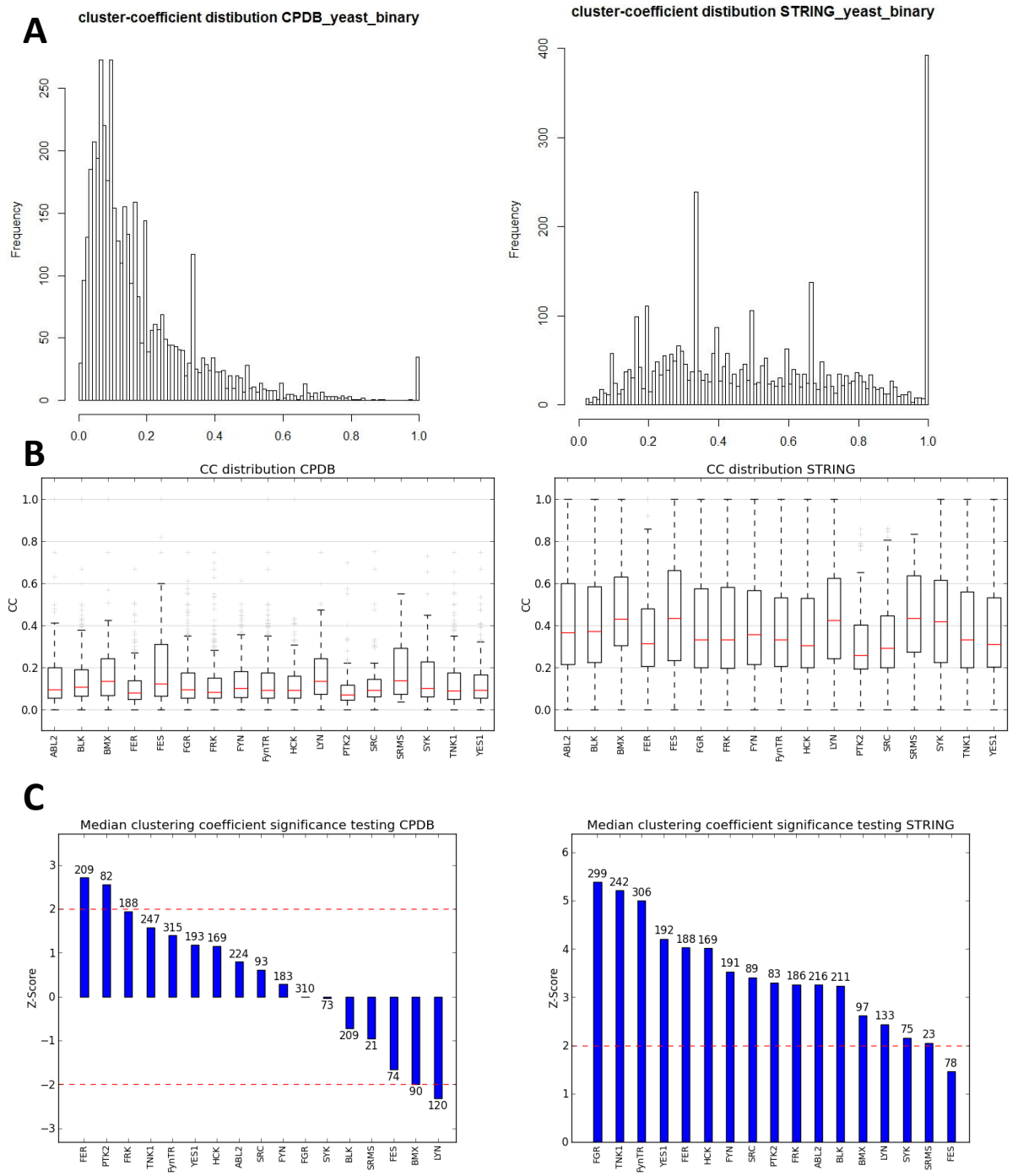


Figure 38: Clustering-coefficient (CC) distributions in the entire networks CPDB and STRING (A) and for each kinase substrate set within the networks (B). Node randomization significance test (Z-score) for each kinase target set in CPDB and STRING (C). The number of targets with CC in the networks written above/below the bars.

In contrast to the kinase targets within the CPDB network, the target sets within the STRING network show significantly higher CCs as expected by random chance and hence cluster in the PPI network

further supporting the notion that cellular processes or protein complexes may be targeted (Figure 38C). Furthermore, other topological measures such as in-betweenness-centrality (IBC) can potentially be exploited to characterize the kinase targets within PPI networks. If a node has high IBC many shortest paths calculated from all pair-wise distances in entire network pass through it suggesting it may represent an important connector node between communities of nodes (Joy et al., 2005, Goh et al., 2002). Overall IBC values distributions followed a tail-like distribution which was a magnitude lower in the CPDB network compared to the more "clustered" STRING network (Figure 39A). Thus, the median IBC values for the kinase target sets was also a magnitude lower (Figure 39B) within the CPDB network while within both PPI networks the first quartile of each target set had an IBC value of zero indicating that at least a quarter of targeted yeast are within clusters of nodes and not in-between cliques of proteins. Network randomization and Z-transformation showed that the median IBC values for most kinase target sets were higher than expected by random chance in the CPDB network, however; lower than expected by random chance for the majority of kinase targets sets in the STRING network (Figure 39C). This result suggests that kinase substrates in yeast are rather in-between network clusters within the CPDB network and rather not in-between, but part of network cliques, within the STRING network.

We also asked whether kinase targets are preferentially interacting. The standard approach to address this question is to shuffle edges between all PPI network nodes while strictly preserving the degree of the nodes and hence the network structure. It was then tested whether the number of interacting pairs within the kinase target node is larger or smaller than in randomized networks. 100 randomizations were performed with an available in-house perl script (Arndt Grossmann and Ulrich Stelzl) using the two yeast PPI networks CDPB and STRING (Table 10). As shown in Figure 40, all kinase target sets form a significantly higher number of interacting pairs than expected by random chance indicated by Z-scores above 2. The fact that rather interacting proteins were targeted suggests models that should involve protein complex targeting or PPI network targeting by the NRTKs, respectively.

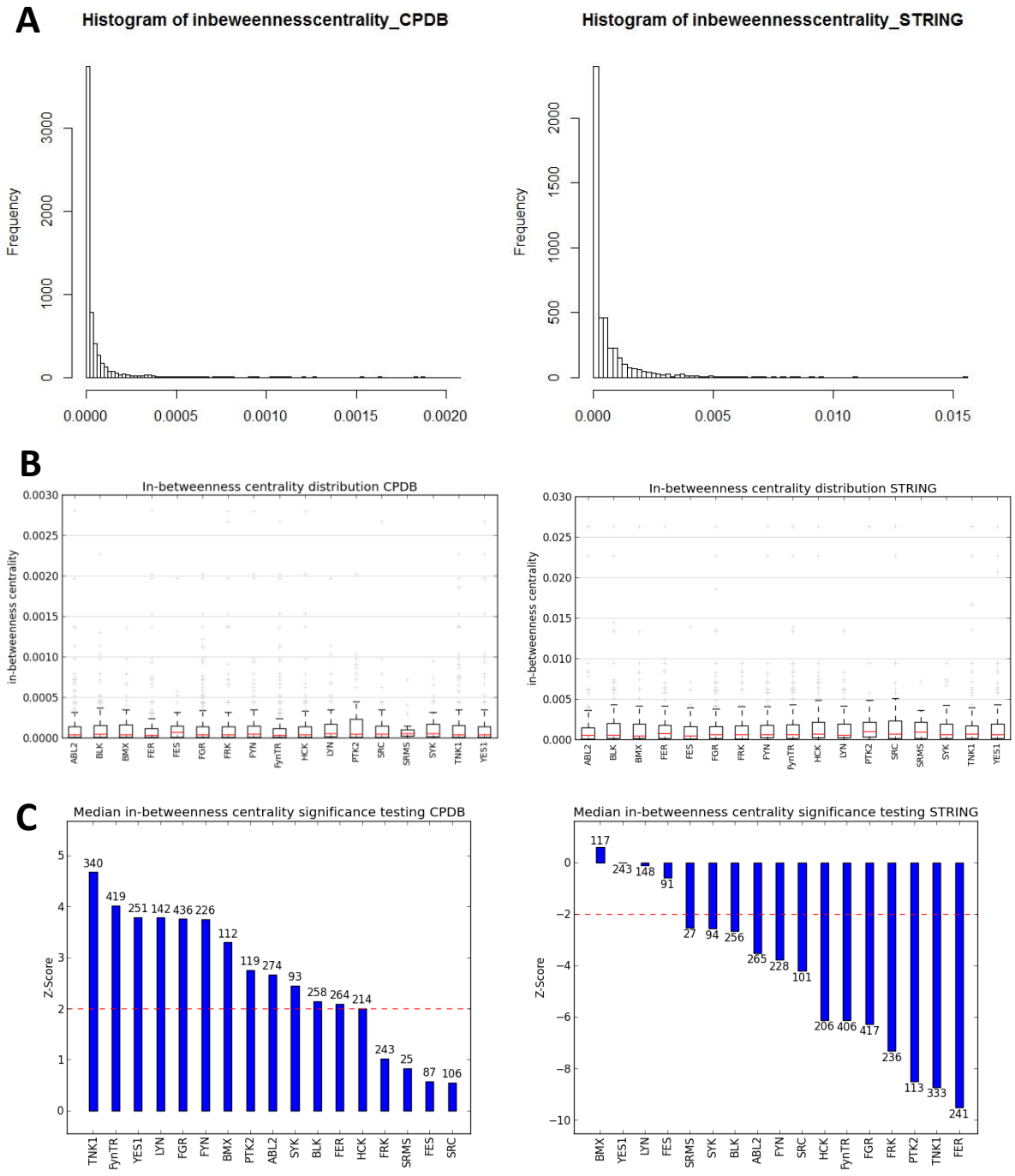


Figure 39: In-betweenness centrality distributions in the entire networks CPDB and STRING (A) and for each kinase substrate set within the networks (B). Node randomization significance test (Z-score) for each kinase target set in CPDB and STRING (C). The number of targets with IBC in the networks written above/below the bars.

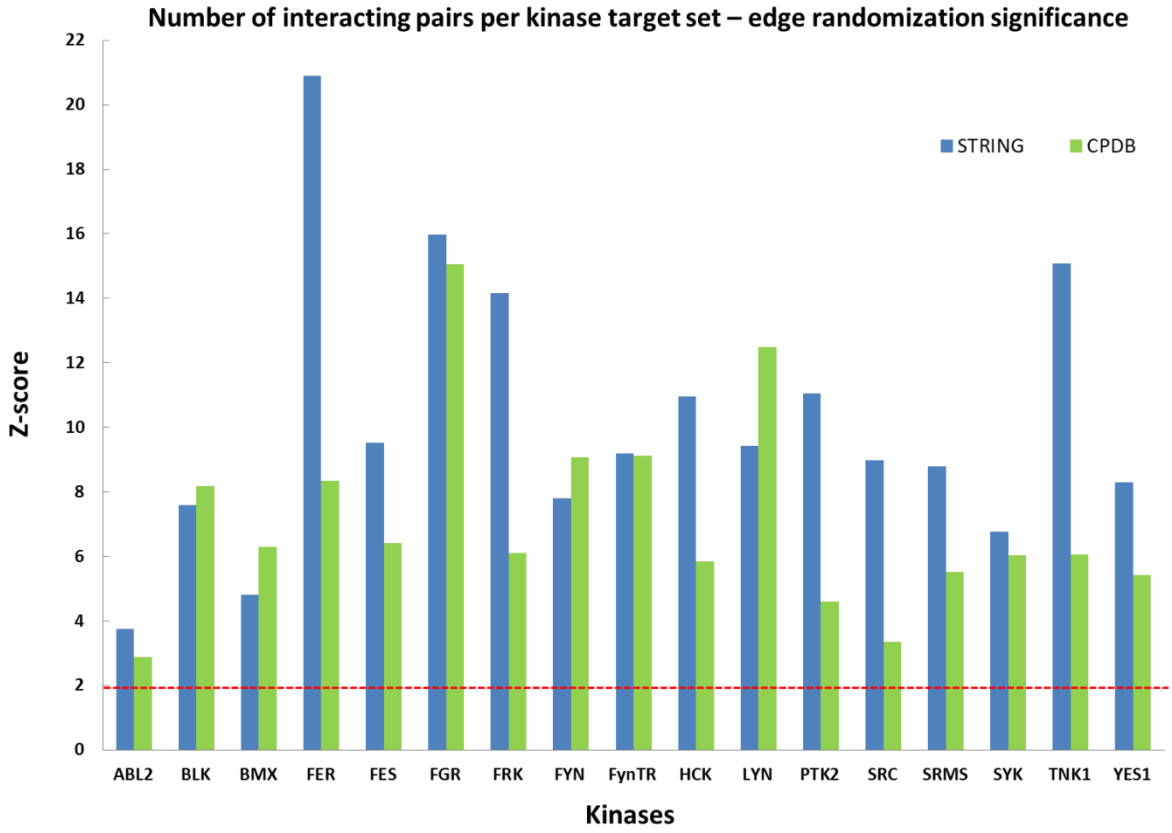


Figure 40: Kinase target set connectivity. By counting the number of interacting target pairs in the yeast networks STRING and CPDB before and after edge randomizations. Z-scores above 2 indicate significantly higher numbers of interacting pairs within kinase targets sets as expected by random chance.

3.16. Kinase targeting in the context of yeast protein complex PPI networks

In human, post-translational modifications including tyrosine phosphorylation selectively accumulate over protein complexes (Woodsmith et al., 2013). Figure 41 A (reproduced from Woodsmith et al. (2013)) shows the number of total pY-sites in human protein complexes normalized to the number of complex components and component size. The red dots represent complexes with accumulated pY-sites distinct from the majority of the data (99% cut-off). We asked whether the tyrosine phosphorylation patterns obtained from our study in yeast results similar signs of accumulation in protein complexes, a yeast protein complex PPI network was obtained from Vinayagam et al. (2013) the analyses was performed using custom-made R scripts (by Dr. Jonathan Woodsmith). Complex components were binned according to the number of complexes they are members of and their size. One hundred randomizations were performed switching complex members preserving the network structure and complex component size distribution. As it can be seen in Figure 41 B, where the median modification levels of complexes was plotted against the total modification levels, a picture

similar to pY-sites in human (Figure 41 A). The measured pY-sites in yeast selectively accumulate over yeast protein complexes. Thus, even though the function of NRTK have not evolved in yeast it appears the yeast cellular machinery was targeted.

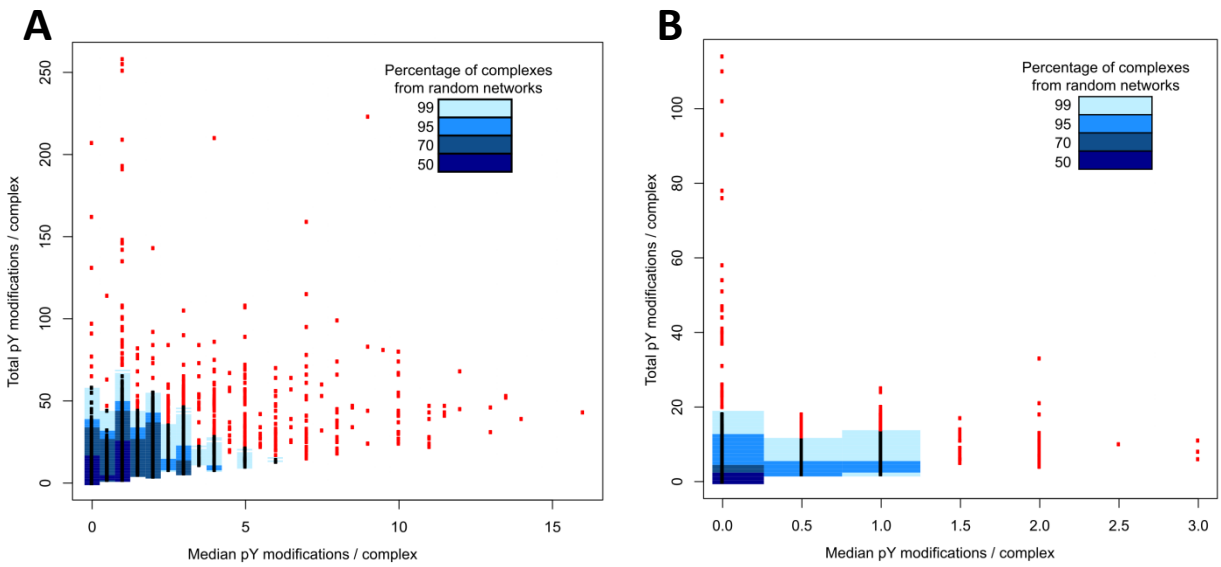


Figure 41: Tyrosine phosphorylation sites accumulation over protein complexes in human (A) and yeast (B). The total number of pY-sites was normalized towards complex component size and degree. Randomization was conducted within frequency bins (in how many complexes a component was present) and size bins to only randomized proteins within the same size quantile. The median number of pY-sites plotted against the total number. Blue background color indicate the percentage of random complexes resulting from node attribute shuffling. Distinct from the majority distribution, the red dots represent complexes that selectively accumulate tyrosine modifications.

4. Discussion

4.1. Yeast is a suitable model system to study NRTK specificity

Saccharomyces cerevisiae (budding yeast) is one of the best characterized model organisms in molecular biology and approximately 30 percent of the yeast proteome is conserved in human (Botstein et al., 1997). For example, proteome-wide protein expression studies have been performed. Ghaemmaghami et al. (2003) conducted global measurements of temporal yeast protein expression. They created a yeast genome-wide ORF fusion library using a high-affinity epitope tag and subsequent proteome-wide enrichment employing a single antibody. Expressed genes were identified using a total of 6234 pairs of synthesized ORF-specific oligonucleotide primers (Ghaemmaghami et al., 2003). Their measurements strongly overlapped with a proteome-wide yeast protein localization study by Huh et al. (2003) who employed green-fluorescent protein (GFP) chromosomal tagging for fluorescence microscopy. In this way, they were able to define subcellular localization for 4156 proteins representing approximately 75 percent of the yeast proteome. The measured of subcellular location correlate well with transcriptional co-regulation and both genetic and physical interactions (Huh et al., 2003). Furthermore, global yeast protein phosphorylation was determined via proteome chips containing about 4400 proteins spotted in duplicate representing *in vitro* substrates recognized by most yeast protein kinases (Ptacek et al., 2005). More recently, a mass spectrometry read-out was used for proteome-wide phosphorylation studies. Quantitative proteomics using stable isotope labeling of amino acids in cell culture (SILAC) enabled comparison between expressed haploid and diploid yeast proteomes (de Godoy et al., 2008). Additionally, a SILAC-based approach enabled sampling of the *in vivo* yeast phospho-proteome in a single experiment comprising 3620, mainly serine and threonine, phosphorylation sites mapped to 1118 proteins using titanium oxide (TiO_2) based phospho-peptide enrichment strategies (Gnad et al., 2009). Furthermore, Breitkreuz et al. (2010) constructed a global yeast kinase-phosphatase interaction network by systematically characterizing protein kinase and phosphatase complexes. Protein complexes were captured via co-immuno-precipitation using magnetic beads, where associated proteins were on-bead digested and identified by mass spectrometry. Thus, a signaling network was constructed comprising in total 201 protein kinases and protein kinase regulatory subunits and 75 protein phosphatases and phosphatase regulatory subunits (Breitkreutz et al., 2010). Moreover, King et al. (2006) created a the *Saccharomyces cerevisiae* PeptideAtlas which was composed from 47 diverse experiments and 4.9 million tandem mass spectra providing the highest degree of proteome coverage for any eukaryotic organism in a single public database. In order to

increase the coverage of protein identification via MS, Picotti et al. (2013) generated a protein sequence dataset based on 6607 protein sequences predicted retrieved from the yeast genome, each one associated with an open reading frame (ORF) in the Saccharomyces Genome Database (SGD, www.yeastgenome.org). Predicted yeast proteins were classified according to their detectability using the PeptideAtlas established by King et al. (2006) and the dataset generated by Ghaemmaghami et al. (2003). For each protein from PeptideAtlas an optimal set of peptides was selected with favorable MS-properties and unique occurrence within the compiled protein sequence dataset. For proteins not included in either PeptideAtlas or the antibody-based dataset likely detectable peptides were predicted and selected using bioinformatics. As a result, approximately 28000 peptide sequences were synthesized to assemble a peptide library representing 97 percent of the predicted yeast proteome (Picotti et al., 2013). By experimentally testing their peptide library for detectability on two types of mass spectrometer set-ups, Picotti et al. (2013) generated two reference spectral libraries supporting discovery-driven (shotgun-MS) and hypothesis-driven (targeted-MS) proteomics. They applied their peptide libraries in proteomic measurements for 78 yeast strains and successfully conducted quantitative trait locus (QTL) analysis which relies on precise measurement of the same set of peptides over a large number of samples. Apart from comprehensive mass spectrometry based studies defining the yeast genome, transcriptome, and proteome, even the yeast lipidome was characterized by quantitative shotgun mass spectrometry (Ejsing et al., 2009). In summary, the yeast proteome is very well defined at all “omics” levels by a plethora of conducted systematic approaches. Thus, yeast is an ideal system for studying phosphorylation specificity on a systems level.

The model organism baker’s yeast is not only well characterized, but also ideal to study human NRTK signaling in a background free manner as no tyrosine kinase orthologs were found and no bona-fide tyrosine kinase activity was detected in fungi (Manning et al., 2002b). When for the first time a cytoplasmic PTK, v-SRC, was expressed in yeast it was observed that yeast growth is diminished and suggested this inhibition of growth is due to kinase activity on yeast proteins (Brugge et al., 1987, Kornbluth et al., 1987). Later reports confirmed that human PTK toxicity in yeast requires an intact, active enzyme and can be explained by aberrant phosphorylation of yeast proteins (Boschelli et al., 1993, Florio et al., 1994). Nevertheless, Nada and colleagues (1991) co-expressed cellular SRC and by the time uncharacterized c-SRC tyrosine kinase (CSK) in yeast cells and discovered that CSK negatively regulates SRC by C-terminal tyrosine phosphorylation. Modular domain dependent SRC regulatory mechanisms were further investigated in *S. cerervisiae* by both Murphy et al. (1993) and Giulio Superti-Furga and Sara Courtneidge, for example (Superti-Furga et al., 1993, Erpel et al., 1995, Koegl et al., 1995, Weijland et al., 1997, Gonfloni et al., 1997, Mandine et al., 2002). Moreover, FES

(Takashima et al., 2003) and HCK (Lerner et al., 2005) regulation was investigated in yeast by co-expression of human kinases and human adaptor molecules. Additionally, c-ABL auto-inhibition was analyzed in *S. pombe* (Pluk et al., 2002) exploiting the absence of (fungal) ABL inhibitory factors. The toxic effect of PTK activity in yeast was exploited for kinase inhibitor screening by searching small molecules or phosphatases which can restore yeast growth by reversing the toxic effect of PTK expression (Montalibet and Kennedy, 2004, Gunde and Barberis, 2005, Koyama et al., 2006, Sekigawa et al., 2010, Harris et al., 2013). In summary, baker's yeast was used previously to study the PTK regulation and biochemical relations between co-expressed human proteins whereas yeast growth phenotypes were observed upon strong expression of PTKs and exploited for PTK inhibitor screening.

PTKs are simultaneously and differently strong expressed in human cells, depending on the tissue (Figure 16), and show magnitude differences in catalytic activity. Due to overlapping substrate specificity it is hence difficult to determine the PTK responsible for *in vivo* measured tyrosine modifications of human substrates. Yeast may be the ideal system to overcome these obstacles. The difficulties in analysis of NRTK specificity in mammalian cells and the principle of using yeast as a model organism is illustrated in Figure 42. In order to gain insight into individual NRTK specificities, previous studies employed randomized linear peptide libraries and hence focused on the determination of optimal linear amino acid sequence motifs (Miller et al., 2008, Mok et al., 2010). However, it was estimated that only around 20 percent of all measured phosphorylation sites match linear sequence motifs (Linding et al., 2007). This indicated that the specificity provided by the substrate protein structure and cellular context needs to be considered (Linding et al., 2007). Duarte et al. (2014) could show by molecular modeling that specificity of measured phosphorylation sites not harboring linear motif information can be encoded in the three-dimensional (3D) structures represented by 3D motifs where critical amino acids can be brought together in space upon folding. In one approach to mimic the expressed human proteome, human cell lysate was treated with proteolytic enzymes providing a database-searchable proteome-derived peptide library in order to identify endoprotease cleavage sites (Schilling and Overall, 2008). Those proteome-derived peptide libraries from trypsin treatment were dephosphorylated using phosphatases to subsequently perform *in vitro* kinase assays for defining kinase specificity and substrates as conducted by Wang et al. (2013). However, the strength of approaches to define NRTK specificity using proteome-derived peptide libraries is still limited by protein digestion thereby likely destroying any 3D fold.

The lack of PTKs in yeast enabled analysis of individual NRTK reactivity in a background-free, highly crowded, eukaryotic environment overcoming previous limitations by displaying fully folded substrates to the enzymes. Due to the lack of regulatory proteins in yeast cells one can hypothesize the absence of NRTK inhibitory and tyrosine modification-dependent mechanisms which may result

in an active state for some of the human NRTKs in yeast. Toxic effects of human NRTK expression in yeast as previously observed for SRC kinases (Superti-Furga et al., 1993) were minimized using a weak expression vector (Figure 9). Expression of human NRTKs in yeast in order to measure their activity on yeast proteins is a novel concept to analyze kinase specificity. This thesis is the first comprehensive analysis of human tyrosine kinase specificity in an intact unicellular eukaryote. During preparation of this thesis, Chou et al. (2012) reported experiments following a similar idea. Two human serine/threonine (S/T) kinases were expressed in *Escherichia coli* and modification changes in the *in vivo* bacterial proteome was measured by mass spectrometry. The author's employed *E. coli* as it has only two kinases with known S/T kinase activity and shows very low levels of endogenous S/T phosphorylation. As a proof of principle, their motifs generated on bacterial sequences mirrored the established kinase consensus sequences and thus, the sequence motifs determined in a lower eukaryote here may even better represent human NRTK specificities.

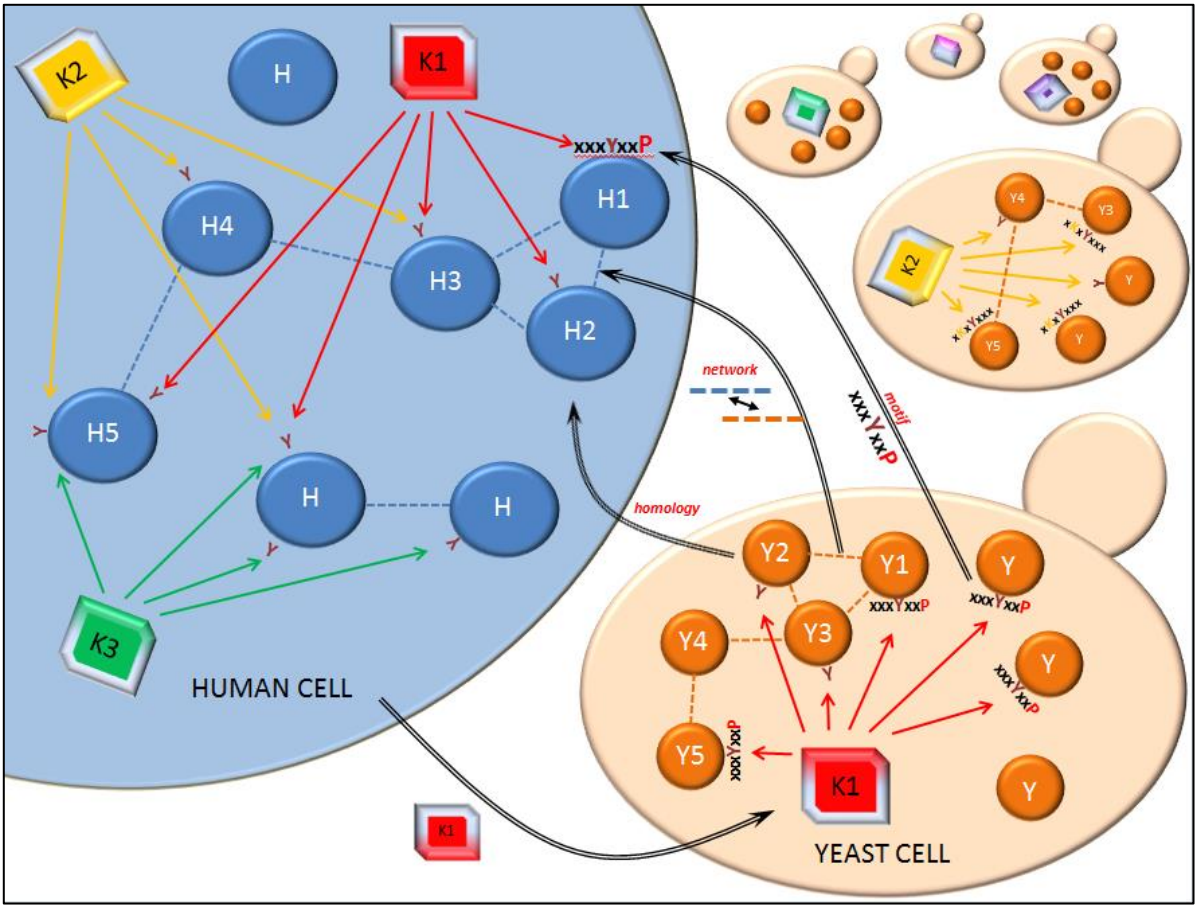


Figure 42: Representation of the concept of using yeast as a model organism for studying protein kinase specificity. In the human cell, kinases K1, K2 and K3 (colored squares) target phosphorylation sites in an overlapping or exclusive manner with strongly differing activity. By expressing an active tyrosine kinase (e.g. K1) in an intact yeast cell lacking NRTKs, the resulting tyrosine phosphorylation of yeast proteins can be measured via MS and attributed to the expressed NRTK. Thus, human KSRs can be inferred from the yeast data via pY-site conservation (homology), obtained specificities in terms of linear sequence motifs, and maybe PPI network information.

In general, human protein kinase specificity can be defined by the combination of linear sequences motif recognition and specificities provided by the cellular environment such as tissue specific expression, subcellular localization, protein complex formation, and protein-protein interactions. The virtually unlimited amount of cellular material from yeast liter cultures and the successful adaptation of previously established immuno-enrichment protocol (Materials and Methods section) yielded large substrates sets. These sets contained only directly targeted pY-sites while protein folding and PPI network structure were preserved in the alive organism. Therefore, KSRs may be transferred from yeast to human by four properties of the yeast target sets: 1) As many yeast proteins have human orthologs one may infer KSRs in human by pY-site conservation. 2) Linear sequence motifs can be derived for each kinase from each individual yeast target set which may be used to predict proteome-wide human modification sites. 3) As the PPI network structure is maintained in intact cells, one may find and exploit conserved yeast PPI network properties such as the formation of protein complexes. 4) Due to the fact that full-length human NRTKs are modifying fully folded yeast proteins, structural properties of NRTK substrates in yeast may be detected and extrapolated to human. In this thesis, human KSRs were inferred by site conservation and linear sequence motifs (Figure 42).

4.2. NRTK substrate specificity is maintained in yeast

The activity of full-length human kinases in yeast varied greatly and NRTK-specific pattern of yeast protein phosphorylation were observed by Western blotting (Figure 8). Crucial regulatory mechanisms such as adapter proteins and activating or inhibiting modification of the NRTKs themselves by other enzymes are, at least in part, missing in yeast. Hence, it is difficult to make assumptions about the extent of enzymatic NRTK activity in mammalian cells based on how many proteins were targeted in yeast. It is possible however, to derive NRTK specificities from the yeast substrates. Recently, Kachroo et al. (2015) replaced 414 essential yeast genes with their human orthologs and assayed for complementation of lethal growth defects upon loss of the yeast genes. Thereby, they could show that nearly half of these yeast genes can be replaced by human counterparts suggesting that these orthologs retained similar or identical functions. Hence, orthologous yeast proteins may be equally well recognized by human NRTKs. Importantly, at least one member of each tyrosine kinase family, assigned according to domain structure, showed activity in yeast enabling a comprehensive analysis spanning an entire major branch of the kinome tree (Figure 7, Manning et al. (2002b)). The observation that over one third of the 900 identified yeast proteins was targeted by a single kinase and approximately two thirds of all proteins were targeted on a single site suggests that targeting specificity of the human NRTKs was maintained in the model

organism (Figure 11 and Figure 12). This assumption is further supported by the finding that roughly half of all targeted sites were modified by only a single kinase and only few sites were targeted by many kinases (Figure 11). Moreover, analysis of protein disorder in targeted yeast proteins revealed no bias of targeted tyrosine residues with respect to protein order (Figure 17). While serine and, to a lesser extent, threonine phosphorylation predominantly occurs in disordered regions a much greater fraction of tyrosine phosphorylation is localized to ordered regions (Iakoucheva et al., 2004). Thus, one can hypothesize yeast tyrosine residues were not arbitrary, but specifically targeted by the human NRTK and were embedded in structurally defined yeast protein surfaces and not primarily on highly disordered structures such as histone tails. Successful expression of at least one member of almost all human NRTK families in yeast enabled comparison of substrate specificity overlap between NRTK families in the same experimental set up as described in the next section.

4.3. NRTK specificity overlap can be detected in yeast

Approximately half of all measured sites in yeast were targeted by a single NRTK which allows identification of individual NRTK-specificity. The other half of measured yeast pY-sites was modified by two or more tyrosine kinases which enables analysis of NRTK specificity overlap in means of shared substrate preferences. The target set similarity between NRTKs was analyzed by hierarchical clustering using Jaccard-indices (Figure 15). Clustering indicated that SFKs shared relatively many phosphorylation sites in yeast. In contrast, FER and FES kinase, representing a separate tyrosine kinase family, appeared to have large differences in specificity. FER clustered with the SRC family kinases and shared many yeast targets with the FRK kinase. FES however, showed relatively little overlap to all other NRTKs analyzed. Moreover, an unexpectedly strong overlap in yeast substrate targeting was observed between kinases having different domain architecture such as for PTK2 (FAK) and TNK1.

In human, the analyzed NRTKs are differentially expressed in tissues (Figure 16). This may partly explain the specificity overlap by assuming NRTKs may fulfill similar functions in different tissues and hence may be under spatial and temporal control as reported for other cytoplasmic kinases (Corson et al., 2003). For instance, FYN and PTK2 are simultaneously expressed in tissues (Figure 16) however, showed very differing substrate specificity (Figure 15). In contrast, PTK2 and TNK1 showed substantial substrate overlap (Figure 15) however, are differentially expressed in human tissues (Figure 16). Linear motif features shared by subsets of kinases were extracted whereas overlapping specificities in yeast proteome targeting can be partly explained by those features (Figure 24). In summary, the comparative analysis of yeast protein targeting for half of all cytoplasmic PTKs within

the same experimental set-up enabled determination of substrate targeting similarity between kinases and kinase families.

4.4. Targeted yeast tyrosine residues are conserved in metazoans

In unikonta, one branch of the taxon eukaryota, metazoa and fungi form the opisthokont branch. Cytoplasmic tyrosine kinases are highly conserved between metazoans, their closest relatives choanoflagellates, and a sister group called filasterea and have likely originated from PTKs of pre-ophisthokont lineages as suggested by Suga et al. (2012). The fungi *Saccharomyces cerevisiae* is the evolutionary closest eukaryote that possesses neither conventional protein tyrosine kinases nor tyrosine-kinase-like enzymes (Manning et al., 2002a). In the cross-species analysis conducted in this thesis, metazoans together with the choanoflagellate *Monosiga bevicollis* formed the group of species having tyrosine signaling evolved (Figure 18). Other eukaryotes such as *Dictyostelium discoideum* have also evolved a much smaller number of PTKs (Tan and Spudich, 1990) however, those earlier evolved PTK-like enzymes were considered to be different from holozoan PTKs as proposed by Goldberg et al. (2006) conducting in-depth analysis of the *Dictyostelium* genome and later by Suga et al. (2012) analyzing genomes of pre-metazoan lineages. At the time of retrieving multiple sequence alignments between yeast PTK targets and orthologous proteins from the INPARANOID database those pre-ophisthokont lineages were placed correctly in the group of species not having tyrosine signaling evolved whereas it was unknown that filasterea have evolved PTKs and hence was not included in the analysis.

For each human NRTK yeast substrate, multiple sequence alignments were retrieved only when a human orthologous sequence was available. Subsequently, it was determined whether the human orthologous sequences contained tyrosine, phenylalanine, or any other residue at the corresponding positions of targeted yeast tyrosine residues. The number of sequence alignments available for each yeast substrate sequence varied as mainly only a subset of all queried species harbored a corresponding orthologous sequence. By normalizing fixed-position tyrosine frequencies to the number of alignments for species having and having not evolved tyrosine signaling, a Y-score and a F-score was generated. The scores enabled analysis of tyrosine and phenylalanine conservation in a single position. Using the Y-score two major cases of tyrosine preservation were observed (Figure 19): On the one hand, a Y-score around zero indicated that the tyrosine (and the majority of surrounding amino acids) was completely conserved throughout all species and suggested the structural necessity of the residues together with a possible functional conservation of the residue. On the other hand, high Y-scores suggested a stronger preservation of the aligned tyrosine residues

in species having evolved PTK signaling opposed to species having not evolved PTK signaling pointing to a functional conservation of that residue. Approximately, half of the measured yeast proteins could be mapped to human orthologs and sets of species having and having not evolved PTK signaling. Over 20 percent of all 310 yeast tyrosine having a human orthologous tyrosine mapped were reported to be phosphorylated in human and hence the corresponding NRTKs were assigned (Table 4). The fraction of conserved tyrosine residues within the group of species having tyrosine signaling evolved was significantly higher when compared to both, all measured pY-sites (all YpY; p=0.0008, Figure 20) and the subset of measured pY-sites with human orthologous tyrosine residues (YpY-HY; p=0.0044, Figure 20) when compared to non-modified yeast tyrosine residues. Thus, measured yeast pY-sites were rather conserved in metazoan lineages suggesting that human NRTK specificity was preserved in yeast cells and that it may be possible to assign kinases to 63 reported human pY-sites by sequence orthology.

A total number of 97 yeast pY-sites was found to be mutated in human to the biochemically most similar amino acid phenylalanine. These sites may be targeted only in yeast as tyrosine was depleted from the metazoan genome during evolution as suggested by Tan et al. (2009b). The researchers detected a negative correlation between species from yeast to human by plotting the percentage of amino acids that are tyrosine against either the number of cell types, or the number of kinases. Tan et al. (2009b) also showed that the fraction of tyrosine is lower in human proteins in respect to their inferred yeast one-to-one orthologs. An example of putative negative selection against tyrosine during evolution is given by Y13 in yeast chaperone Stress-Seventy subfamily A (SSA1) and the corresponding site in human chaperone HSP8 in Figure 20. The multiple sequence alignment shows the mutation of tyrosine to phenylalanine almost exclusively within the TK group. Apart from *Strongylocentrotus purpuratus* only the three members of the TK group evolutionary most distant to human have a preserved tyrosine at this position. Interestingly, four species reported to have pre-ophisthokont tyrosine kinases also have the Y13F mutation. Promiscuous tyrosine phosphorylation changes protein function and stability, and can be lethal as shown by strong PTK expression in yeast (Superti-Furga et al., 1993). Tan et al. (2009b) proposed a positive selection of tyrosine loss in order to prevent promiscuous tyrosine modification. However, recently Pandya et al. (2015) followed up on this hypothesis by statistical testing of predictions based on the assumption that deleterious phosphorylation was driving tyrosine loss in metazoans. Their results indicated that tyrosine loss in metazoans cannot be explained by increased tyrosine phosphorylation propensity due to solvent accessibility, structural disorder, protein abundance, and mutational frequencies for creating or deleting tyrosine when compared to other amino acids and between yeast and human. Thus, their findings contradicted the promiscuous phosphorylation hypothesis of Tan et al. (2009b) and hence it was suggested that tyrosine loss is neutral with respect to phosphorylation (Pandya et al., 2015). It

remained unclear whether other selection mechanism contributed to the loss. Independent of the overall tyrosine residue loss in metazoans, human NRTK targeted sites in yeast are conserved in metazoans and a subset of tyrosine targeted in yeast was mutated to phenylalanine in metazoans. This may suggest negative selection of specific tyrosine residues for further examination - maybe as a result of fine-tuning NRTK signaling during evolutionary expansion of the three tier tyrosine signaling repertoire.

4.5. Linear amino acid sequence motifs including contextual information were generated

NRTK specificity can be defined by kinase domain recognition of the amino acids framing the modification site. The published linear motif atlas by Miller et al. (2008) is a benchmark study providing amino acid specificities based on optimal peptide sequence binding the kinase domain for approximately 35 percent of the human kinome including five cytoplasmic PTKs. From each measured kinase target set in yeast novel and improved linear sequence motifs were derived from expressed, fully-folded, and interacting yeast proteins. In order to retrieve motifs with maximum information content at least four amino acids positions to both sides of the modification site have to be considered as determined over twenty years ago: Songyang et al. (1994) conducted crystallographic analysis and peptide screening and Kreegipuu et al. (1998) performed statistical analysis on around 1000 reported pY-sites. Positive and negative enrichment of residues on seven positions flanking the targeted pY-site in yeast was determined for each kinase target set using custom-made PYTHON programming language based scripts (material and methods 2.6.1) and the iceLogo application (Colaert et al., 2009). The analysis of 15mers may yield additional information as more recently shown by computing the profile matrices for several hundred kinases (Safaei et al., 2011). Further extension of this sequence frame, in particular the analysis of 15 or more positions adjacent to the pY-site may lead to increased noise in training sets and hence may complicate kinase-substrate predictions (Safaei et al., 2011). For most cytoplasmic PTKs still no linear phosphorylation motif has been determined to date (Miller et al., 2008) and novel motifs as for TNK1, for instance, could be retrieved by the approach presented here. Furthermore, crucial elements of literature described linear motifs could be reproduced in yeast and additionally, previously unknown specificity determinants were identified. For example, screening of v-ABL on proteomic libraries derived from cellular mRNA identified ABL family kinase domain preference for proline at position +3 C-terminal and for aliphatic amino acids at position -1 (Cujec et al., 2002). Querying all known ABL1, ABL2, and BCR-ABL1 substrates for amino acid enrichments Colicelli (2010) determined an additional

preference for negatively charged amino acids at positions -3, -4, and +1 flanking the modification site as shown in Figure 43 where the yellow bar indicates the position of the tyrosine.

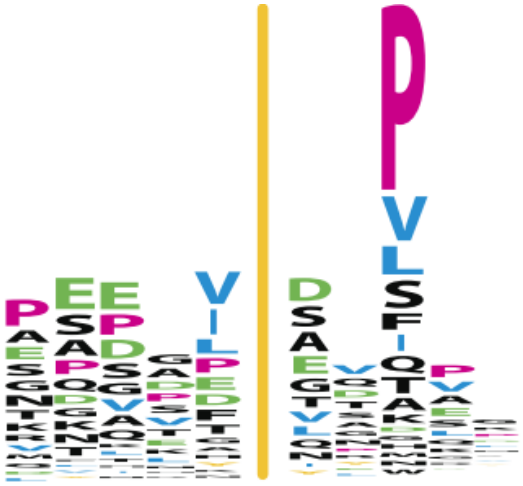


Figure 43: ABL kinase consensus sequence motif taken from Colicelli (2010).

The ABL2 motif derived from yeast substrates recapitulates these preferences (Figure 23). Additionally, previous knowledge was extended by determination of ABL2 preference for a stretch of positively charged residues at positions +4 to +7 and the previously neglected under-representation of both, negatively charged residues within this C-terminal stretch and positively charged residues on positions -1 and -2 (Figure 23). In summary, novel and well performing linear motifs were derived from yeast preparations - recapitulating and extending previously defined substrate targeting determinants.

Linear amino acid sequence motifs (herein after referred to as “motifs”) derived from folded yeast protein structures may perform better in determining sequence specific phosphorylation sites than motifs generated by linear peptide library screening. Structural information in the vicinity of the modification sites in yeast may be partly embedded within retrieved primary sequences. The linear motif atlas (Miller et al., 2008) was constructed with the prerequisite that every motif was based on at least 12 reported pY-sites after selection and hence sites were grouped by family for tyrosine kinases. However, family grouped motifs are mixtures of single kinase motifs which may be significantly different from each other and hence may be unfeasible for inference of kinase-specific substrates. The problem of kinase family motifs masking individual kinase motifs was recently shown in a study concerning the impact of genetic variation on substrate targeting by kinases (Wagih et al., 2015). Wagih and colleagues (2015) retrieved 7004 kinase-associated phosphorylation sites from public databases and constructed a total of 476 PSSM-based sequence motifs including 58 kinase

family motifs of 322 kinases with varying confidence. Interestingly, the researchers refined their motifs by iteratively discarding sequences with poor correspondence to the PSSM similar to the refinement of motifs conducted in the approach presented here (Wagih et al., 2015). The generated motifs for cytoplasmic PTKs in their approach are very different from the motifs obtained here from yeast data (Figure 44). This may be due to the limited number of known kinase-substrate associations in databases. There were seven motifs generated for NRTKs (ABL2, BMX, FYN, LYN, SRC, SYK, and YES1) which were also included the yeast experiments as shown in Figure 44. The amino acid frequencies of ABL2 targets were based on only 14 known substrate sequences (as compared to a total of 280 pY-sites obtained from ABL2 expression in yeast) which enabled solely identification of enriched valine adjacent N-terminal (position -1) to the substrate tyrosine and enriched proline three amino acids C-terminal (position +3) to the substrate tyrosine (Wagih et al., 2015). Nevertheless, the enrichments are in agreement with the yeast-derived ABL2 specificities. The BMX motif of Wagih et al. (2015) shows no similarity to the BMX motif generated in yeast however, both sequence motifs for FYN show similar amino acid enrichments at positions -3 (D/E), -1(V/L), +1 (D), and +3 (V/L/I). Furthermore, both LYN sequence motifs indicate a preference for glutamate at position +1 and leucine at position +3. The main specificity determinant provided by Wagih et al. (2015) for SRC substrates is a negatively charged residue (E/D) at position -3. Strikingly, the yeast-derived SRC motif delineates negatively charged residues (E/D) at position -3 as well.

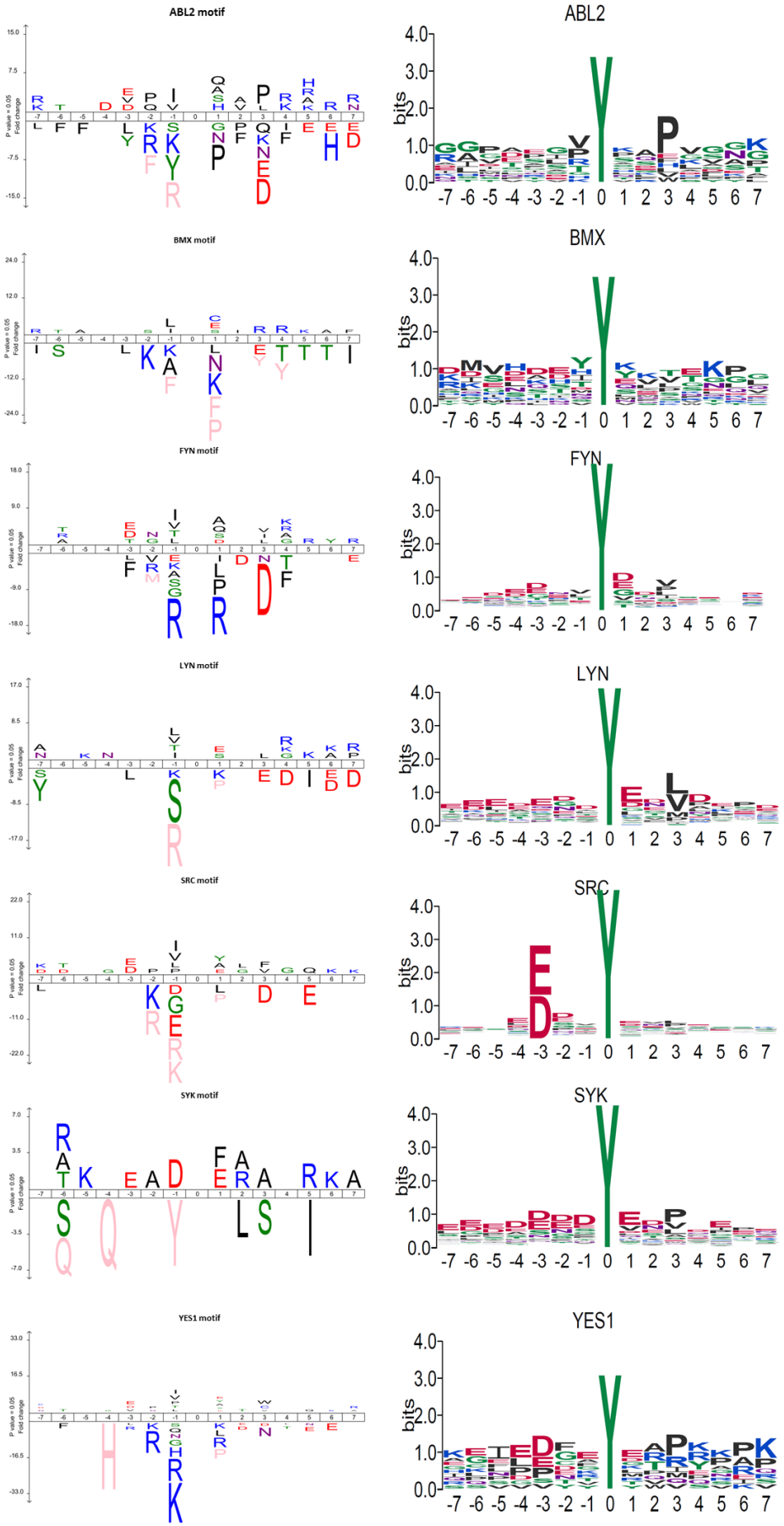


Figure 44: Comparison of motifs determined in yeast (left side) to the motifs generated by Wagih et al. (2015) (right side). Explanation in text.

The yeast-derived SRC motif suggests however, that also other over- and under-represented residues may contribute to SRC substrate specificity. Similarly, both YES1 motifs harbor a negatively charged residue (E/D) at position -3, enriched glutamate at position +1, and threonine at position +2. Finally, the SYK motif of Wagih et al. (2015) shows enrichment of solely negatively charged residues N-terminal and at two positions C-terminal to the modified tyrosine residue. Similarly, the yeast-derived SYK motif shows enrichment of negatively charged residues at positions -3 and -1 and an over-representation of glutamate directly C-terminal to the tyrosine residue (Figure 44). Thus, the yeast-derived sequence motifs recapitulate in part the motifs generated by Wagih et al. (2015). As the latter approach is limited to the available number of known kinase substrates in databases, the motifs generated experimentally employing a greater number of yeast substrates may be more precise representations of NRTK specificities.

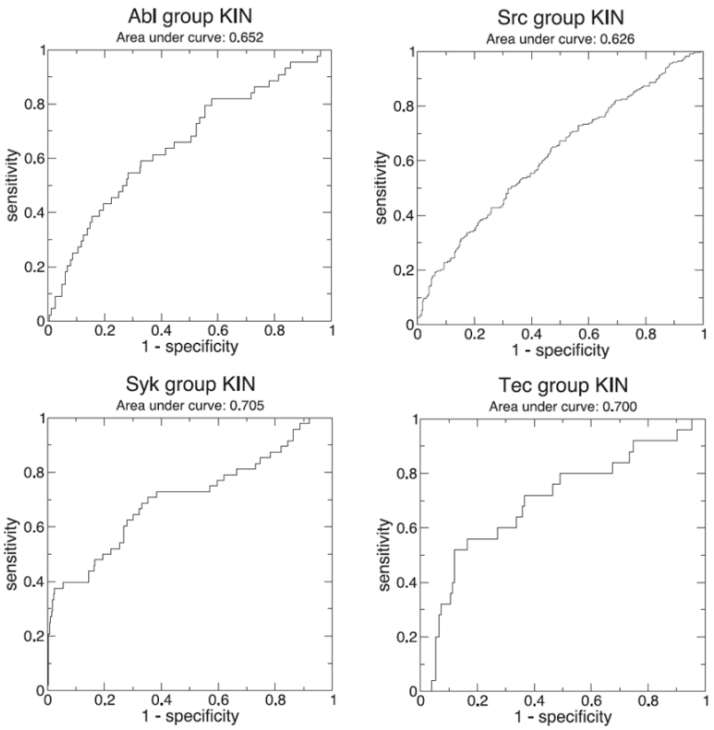


Figure 45: ROC plots for the ABL group, SRC group, SYK group, and Tec group motifs mainly based on linear peptide screening and optimized by phylogenetic comparisons taken from Miller et al. (2008).

For performance comparison, ROC performance graphs for NRTK group motifs generated by Miller and co-workers (2008) for which the kinases are included in the yeast experiments as well are shown in Figure 45. The SRC-group, TEC-group, and SYK-group motifs are hardly comparable to the individual motifs for each SRC-kinase, the motif of the only analyzed member of the Tec-kinase family BMX, and the SYK motif which was generated independent of ZAP70 specificities, respectively.

Individual NRTKs motifs performed both better and worse than the group motifs of Miller et al. (2008) whereby comparing performance to average AUCs over kinase families may be inadequate. The ABL group motif provided by the linear motif atlas may be most comparable to the ABL2 motif presented here. ABL1 and ABL2 kinase domains share 90 percent sequence identity and multiple sequence alignments showed that both kinases are more similar to each other than any other non-receptor PTK (Colicelli, 2010). Also the ABL1 and ABL2 motifs generated by Wagih et al. (2015) are highly similar. Miller et al. (2008) determined an area-under-curve (AUC) value of 0.652 for their published ABL group motif generated by linear peptide screening which is much lower than the AUC of 0.77 for the ABL2 motif obtained from the yeast proteome as determined by cross-validation (Figure 25B).

The complete absence of certain residues surrounding the measured yeast pY-sites suggests that occurrence of non-permissive amino acids in substrate sequences is a prevalent substrate specificity determinant (Liu et al., 2010). Structural constrains on phosphorylation site residue frequencies was suggested already a decade ago by Blom et al. (1999) who constructed the first sequence and structure based phosphorylation site predictor using neural networks (NetPhosK). By comparing amino acid frequencies of 210 pY-sites reported in total, the authors observed that tryptophan, cysteine, and methionine never appeared at specific positions flanking the tyrosine and proposed structural accessibility as the origin of their absence. In addition to the experimental data presented here, this assumption is further supported by the findings of Alexander and colleagues (2011) who performed molecular modeling on X-ray crystallography structures of mitotic kinases together with linear peptide screening. By 3D-modeling, the importance of evolutionarily conserved motif exclusivity, i.e. non-permissive amino acids within linear motifs together with NRTK subcellular localization, for kinase specificity was demonstrated. Furthermore, considering evolutionary aspects of NRTK targeting by employing their NetPhorest algorithm (Miller et al., 2008), Tan et al. (2009b) predicted that human tyrosine residues are rather not present within known linear motifs. This further supports the notion that pY-site surrounding amino acids may evolve in part also to prevent unfavorable phosphorylation.

Even though motifs generated from yeast target sets appear to perform better than previously established versions, the FDR in predicting the previously measured kinase targets in yeast was high even when applying highest possible accuracy cutoffs (Figure 26A). Assuming that also in yeast only up to 20 percent of targeted tyrosine reside within motifs (Linding et al., 2007), the best scoring 20 percent of pY-sites from each yeast kinase target set were used to redraw the motifs. By using the refined motif and the determined optimal high accuracy cutoff at 0.995 for phosphorylation site scoring, a higher number of sites was predicted with lower FDR (=0.278) when compared to the

original motif (FDR=0.344) (Figure 26B). Occurrence of pY-site flanking non-permissive residues is strongly emphasized in the refined motifs as several amino acids never appear within the best scoring sites (Figure 27). Due to the smaller number of sites for generating the refined motifs it was not feasible to conduct cross-validation for refined motif performance analysis. In direct comparison to the original motif however, the refined motif performed better in retrieving the 20 percent best scoring yeast pY-sites from the yeast proteome (Figure 28, inset) and worse at higher sensitivity, i.e. in finding possibly more of the measured kinase targets (Figure 28). Thus, the refined motifs were chosen for NRTK target site predictions in human.

4.6. NRTK yeast target sets cluster in PPI-networks

Due to the fact that PTK signaling has not evolved in *Saccharomyces cerevisiae* it was possible to analyze the influence of PPI network structure on substrate targeting for individual NRTKs. When analyzing PPI network properties for each kinase target set individually it was detected that NRTK-modified yeast proteins were much closer within interaction networks than expected by random chance. The average shortest path between proteins phosphorylated by a kinase was significantly shorter in both yeast networks analyzed, i.e. a PPI network obtained from STRING and a binary network obtained from CPDB (Figure 37). Together with the assumption that only 20 percent of substrates can be recognized by NRTKs in a modification sequence dependent manner, significant network proximity of targeted nodes hints towards substrate targeting influenced by PPI network context as outlined in Figure 46.

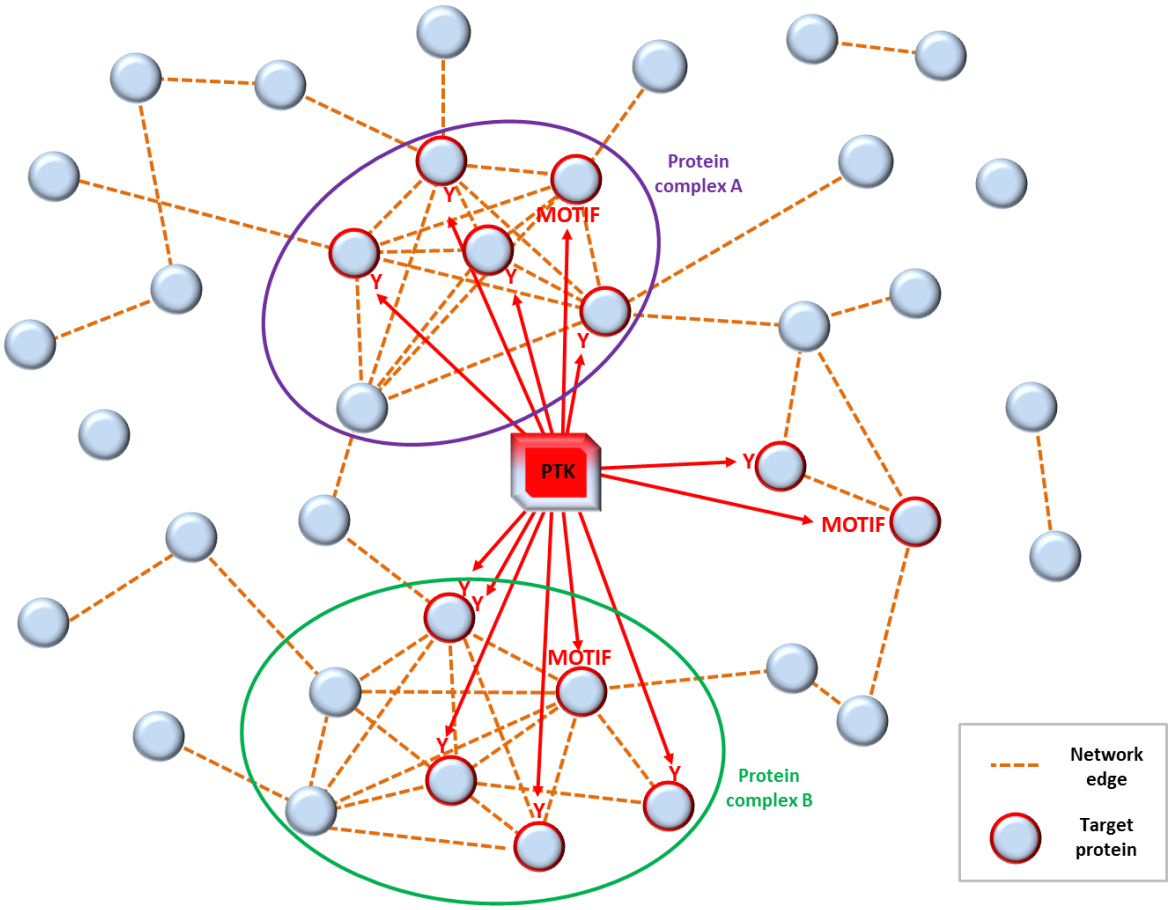


Figure 46: Schematic representation of PPI network directed NRTK targeting. Proteins/nodes are shown as light blue spheres, edges as dashed orange lines, targeted nodes are circled in red, protein complexes are circled in violet and green, and substrates harboring motifs are labeled.

It has to be considered that the network topology of the physical, binary yeast PPI network retrieved from CPDB is very different from the dense, binary, probabilistic yeast PPI network from STRING database using a STRING confidence score high cut-off. Kinase target sets within the STRING network have a much higher average median CC than within the CPDB network (Figure 38B). When comparing the CCs of each kinase target set to all other nodes within each network for significance testing, it was observed that, in contrast to target sets within the CPDB network, target sets within the STRING network were more inter-linked than expected by random chance (Figure 38C). Thus, it can be suggested for the STRING network that protein complexes or sub-networks of highly interacting nodes were targeted. In regard to the third investigated network property in-betweenness centrality, it was detected that kinase target set nodes within the CPDB network tend to have high in-betweenness centrality which may indicate that nodes linking cliques of network nodes or sub-networks were targeted (Figure 39C). However, this was not observed for yeast target sets within the STRING PPI network (Figure 39C). The notion that interacting pairs and protein complexes were

targeted in yeast was further supported by counting the number of interacting pairs of nodes within target sets in comparison to a randomized network with preserved topology (Figure 40).

The discovery that tyrosine phosphorylation sites accumulate over yeast protein complexes, equivalent to human tyrosine modifications (Woodsmith et al., 2013), further points towards NRTK targeting of PPI networks (Figure 41). The spatial accumulation of pY-sites on protein complexes may result from fast rebinding of NRTK's modular domains to the complex after recruitment to pY-sites by primary sequence information, for instance, provided by one or few protein complex members. This was also suggested by Oh et al. (2012) for SH2 binding proteins in protein tyrosine kinase receptor signaling. Oh et al. (2012) used a total internal reflection (TIR) microscope to measure single SH2 molecule movement from its initial docking pY-site at the cellular plasma membrane out of the TIR illumination field. The authors demonstrates that the average dwell time was longer than predicted on the basis of the SH2 domain chemical dissociation rate supporting a model of fast SH2 domain rebinding. Thus, NRTKs may modify tyrosine in close proximity within the PPI network with specificity provided by the spatial organization in complexes of putative substrates and increased dwell time by rebinding to priming phosphorylation events via modular domains. It may also be possible that NRTKs do not have to re-bind via their interaction domains, but may also be able to modify tyrosine residues in temporal and spatial proximity by chance.

In summary, by performing statistical testing on yeast PPI network targeting it was disclosed that human NRTKs may be guided by the PPI network to approach potential tyrosine modification sites. Moreover, the results may suggest that substrate specificity for NRTK targeting is additionally provided by complex assemblies, i.e. the spatial and temporal coordination of kinase and substrate.

4.7. Structural analysis points towards coordination of substrate targeting via protein complex assemblies

The analysis of yeast substrate targeting within PPI network suggested targeting of highly interacting proteins or macromolecular protein assemblies. Thus, tyrosine phosphorylation sites were mapped to solved, 3D yeast protein complex structures. In Figure 47, all surface exposed tyrosine residues were mapped on the structure of the yeast 40S ribosomal subunit and labeled in red whereby measured, targeted pY-sites were marked in yellow. It can be observed that the phosphorylation sites are not randomly distributed over the protein complex, but in relative proximity on one side of the protein complex. The fact that there are no modification sites mapped to the interior of the complex may suggest post-assembly complex targeting by NRTKs. Finally, the discovered local

accumulation of pY-sites on different protein complex subunits may further indicate that substrates may be presented to the kinases via the spatial organization of protein complexes.

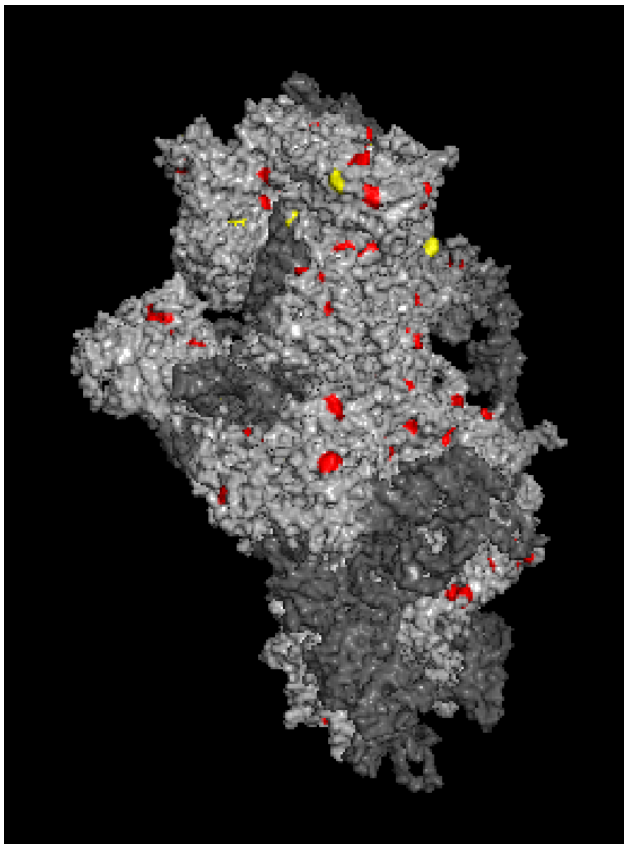


Figure 47: Structure of the 40S ribosomal subunit with all surface exposed tyrosine residues in red and targeted tyrosine residues in yellow. (Mapping and figure by Sean R. Connell (JWGU, Frankfurt)).

4.8. Kinase specificity learned from yeast can be directly transferred to human

The importance of contextual factors in PTK signaling and in particular in the prediction of KSRs was previously shown (Blom et al., 1999, Linding et al., 2007, Newman et al., 2013, Damle and Mohanty, 2014, Patrick et al., 2015). Linding et al. (2007) could double linear motif based prediction accuracy by employing a contextual network constructed from the probabilistic interaction resource STRING (Snel et al., 2000). The developed NetworKIN algorithm (Linding et al., 2007) employs the NetPhorest algorithm (Miller et al., 2008) for motif generation and kinase assignment. In brief, reported pY-sites and protein domain sequences were retrieved from public databases, and kinase specific PSSMs from *in vitro*, peptide library-based kinase assays. The data for each kinase was subsequently organized and filtered via phylogeny, and used as training set for classifier determination using artificial neural networks. Finally, STRING was employed to build a context network for each substrate based on

interaction and pathway databases, literature mining, studies on mRNA expression and genomic co-occurrence and kinase-substrate relationships were inferred. The NetworKIN methodology is summarized in Figure 48 and compared to the approach presented here.

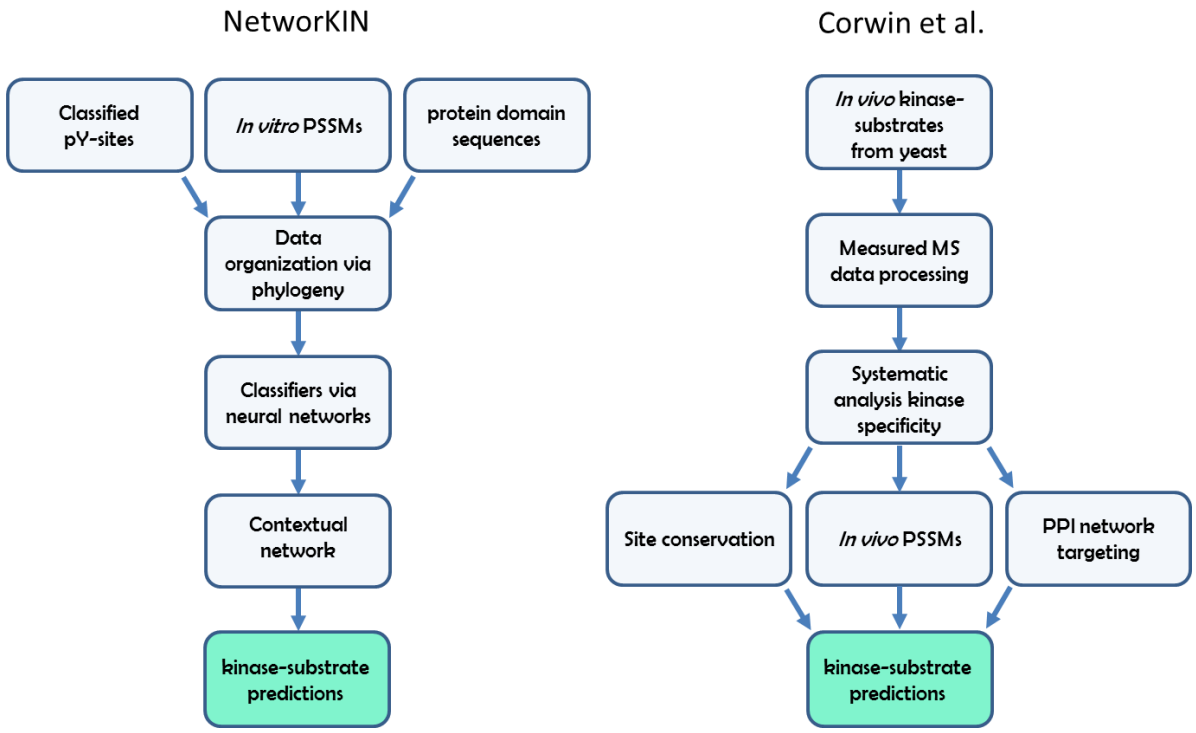


Figure 48: Schematic representation of workflow for kinase-substrate assignment by the NetworKIN algorithm and by the novel approach using baker's yeast as a model organism.

NetworKIN relies on the limited number of reported KSRs which are prone to investigational biases and *in-vitro* inferred specificities for predictor generation. The approach presented in this thesis has the advantage of testing each NRTK individually in a single experimental set-up where large sets of targeted modification sites can be retrieved for the investigation of kinase specificities. Using the same expressed, intact yeast proteome for the analysis for all NRTK enabled inter-kinase comparisons. Furthermore, the experimental system mimics a highly crowded cellular environment with preserved (yeast-specific) contextual factors and fully folded substrates in a living eukaryotic organism. Besides structural aspects considered in yeast, a more natural effect size for generating NRTK sequence motifs via PSSMs is provided when compared to infinite sequence availability in randomized peptide screening. Thus, those PSSMs generated quasi "*in vivo*" may perform better than their *in vitro* counterparts and may enable more accurate kinases inference for reported pY-sites. Moreover, the fact that approximately 30 percent of the yeast proteome is conserved in human (Botstein et al., 1997) enables direct NRTK prediction for conserved human pY-sites in a phylogenic analysis. NetPhorest used kinome domain phylogeny as an additional proxy for NRTK targeting

similarity for site annotation. This is entirely different from the cross-species analysis for site conservation from yeast to human conducted here. Another benefit of the established NRTK assay here is the possibility to study the direct activity of each individually expressed NRTK on physical and probabilistic PPI networks. Thus, it may be possible to boost prediction of KSRs in human by including network property extrapolation. The results from exploiting the advantages of using yeast as a model system are discussed in the next paragraphs.

4.8.1. Site conservation between yeast and human enables kinase inference and suggests extensive NRTK signaling in glycolysis

4.8.1.1. The Warburg effect

Oncogenic PTK signaling can have great influence on cellular respiration and may lead to aerobic glycolysis also known as the “Warburg effect” (Vander Heiden et al., 2009). Already in the 1930s, Otto Warburg observed that cancer cells show “damaged respiration” and compensate for energy shortage not by oxidative phosphorylation in the mitochondria, but via less effective anaerobic lactate production via glycolysis - even with sufficiently supplied oxygen (Warburg, 1956). Warburg proposed a defect in the mitochondria as the origin of respiration chain disturbance. However, more recent investigations showed that mitochondria are intact in proliferating cells and that levels of energy-carrying molecules adenosine triphosphate (ATP) and the reduced form of nicotinamide adenine dinucleotide (NADH) are equally high compared to normal cells which led to other explanations as reviewed by Vander Heiden et al. (2009). The most conclusive explanation reported by Vander Heiden et al. (2009) is based on the assumption that proliferating cells such as cancer cells, compared to normal cells, have different metabolic requirements for glucose and glutamine, and glycolytic intermediates. These metabolites can be used as carbon and nitrogen source for synthesis of fatty acids, amino acids and nucleotides (Vander Heiden et al., 2009, Mazurek, 2011). Furthermore, elimination of excess carbon via increased lactate excretion can lead to faster production of biomass during proliferation (Vander Heiden et al., 2009). By assembling data from over 2500 human gene expression microarrays spanning 22 different tumor types and corresponding healthy tissues, Hu et al. (2013) showed that changes in metabolic enzyme expression is a common property of tumors. This argument was supported by a strong correlation between tumor samples and comparisons between experimental investigations and set-ups. Furthermore, the authors determined that metabolic processes were up-regulated in tumors that are important for anabolic biosynthesis including the pentose phosphate pathway, glycolysis, and purine and pyrimidine synthesis. On the other hand, growth inhibiting processes such as fatty acid degradation and

essential amino acid degradation were down-regulated in tumors. Their most interesting finding however, was the discovery of heterogenic expression of all metabolic enzymes, and in particular of those involved in oxidative phosphorylation and the citric acid cycle, among different tumors. Therefore, it was suggested that up- or down- regulation of genes encoding tricarboxylic acid (TCA) cycle enzymes dependent on tumor tissue is an adaption to the tumors' specific needs and not a common feature in cancer such as increased expression of glycolytic enzymes in the vast majority of tumors. Moreover, there is growing evidence that the switch to aerobic glycolysis in tumors and leukemia is a result of cross-talk between signaling pathways and metabolic control to adapt for the anabolic requirements of proliferating cells (Vander Heiden et al., 2009). To a great extent, this cross-talk and co-ordination between glycolysis and anabolic biosynthesis is facilitated by changing metabolic enzyme activities via post-translational modifications (PTMs) including lysine acetylation, glycosylation, cysteine oxidation, and phosphorylation as reviewed by Hitosugi and Chen (2014).

4.8.1.2. Site conservation identifies glycolytic enzymes as NRTK substrates

By mapping yeast pY-sites to human orthologs, it was discovered that a total of 63 measured sites in yeast have an orthologous tyrosine residue in human reported to be phosphorylated (Table 4). 13 of these orthologous human pY-sites were predicted both by homology and by sequence motif to be targeted by some NRTKs (Table 7). For nine of these pY-sites some NRTKs were inferred by both prediction methods providing additional confidence in kinases assignment. The fact that several kinases can be confidently assigned to the same pY-sites indicate the strength of the experimental set-up to predict and identify NRTK specificity overlap.

Gene ontology (GO) term enrichment analysis concerning the 63 conserved pY-sites showed that human NRTKs were targeting yeast proteins involved in conserved cellular processes such as translation and metabolism. Strikingly, all proteins of the enriched glycolytic pathway were targeted by 13 out of 16 NRTKs screened as highlighted in Table 4 and Figure 22. Interestingly, over 80 percent of conserved yeast enzymes in carbohydrate metabolism may be substituted by human orthologs (Kachroo et al., 2015). A total of 14 of the 63 orthologous sites targeted in yeast, approximately 22 percent, reside on eight metabolic enzymes involved directly in glycolysis or are involved in cellular energy turnover as summarized in Figure 49. Apart from fumarase and malate dehydrogenase, all metabolic enzymes involved in the citric acid cycle were targeted by human NRTKs in yeast. In contrast to the glycolytic enzymes however, not a single modified tyrosine of yeast TCA-enzymes was reported to be phosphorylated in human orthologs. Therefore, the glycolytic pathway appears as a hot spot of conserved NRTK-targeted sites of metabolic enzymes. Glycolytic enzymes form, together with further metabolic enzymes and signaling molecules, the "glycolytic enzyme complex" (GEC) as

reviewed by Mazurek (2011). NRTK targets in yeast involved in glycolysis with conserved pY-sites in human are shown in Figure 49. Protein interactions from CDPB and STING binary, yeast PPI networks were mapped to the metabolic enzymes to indicate the physical and probabilistic connectivity with the GEC.

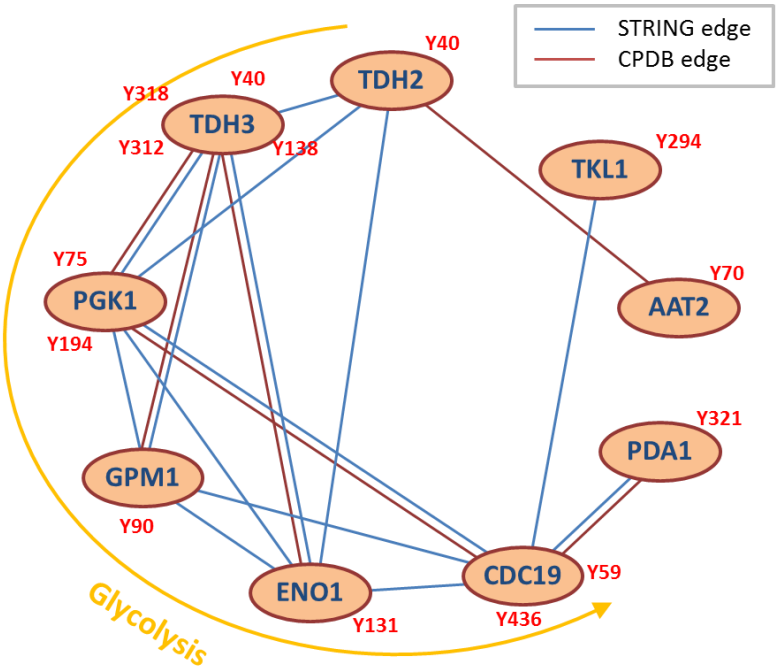


Figure 49: Simplified schematic view of PPIs among targeted yeast metabolic enzymes involved in sugar metabolism created using STRING (blue edge) and CPDB (dark red edges) yeast PPI networks. Targeted residues indicated in red. The yellow arrow indicates direction of metabolite conversion by the "glycolytic enzyme complex".

Targeting of glycolytic enzymes was possibly guided by supra-molecular organization. The fact that none of the conserved targeted pY-sites match NRTK motifs further suggests protein complex guided NRTK targeting. All human metabolic NRTK substrates predicted via sequence similarity to targeted yeast proteins are presented in a schematic of glucose metabolism (Figure 50). Potential regulatory roles of cytoplasmic PTK signaling in glycolysis are discussed as a demonstration for the feasibility of direct NRTK specificity transfer from yeast to human via site conservation. Two predicted KSRs were experimentally verified: targeting of Y92 of phosphoglycerate mutase (PGAM) by ABL2 and targeting of Y76 of phosphoglycerate kinase 1 (PGK1) by FGR. Starting with the intermediate metabolite glyceraldehyde-3P, every metabolic enzyme involved in subsequent conversion steps to the final product of glycolysis, pyruvate, was targeted in yeast on tyrosine residues by several tested NRTKs, but not SRC and LYN. These tyrosine residues were reported to be modified in human as well. Notably, FER was the NRTK which targeted all metabolic enzymes in yeast involved in converting glyceraldehyde-3P to pyruvate and was moreover the only kinase showing activity towards the

majority of metabolic enzymes within the TCA cycle (as discussed above). Similarly, FGR modified most glycolytic enzymes and hence it may be speculated that FGR and in particular FER may be heavily involved in controlling glycolysis in human.

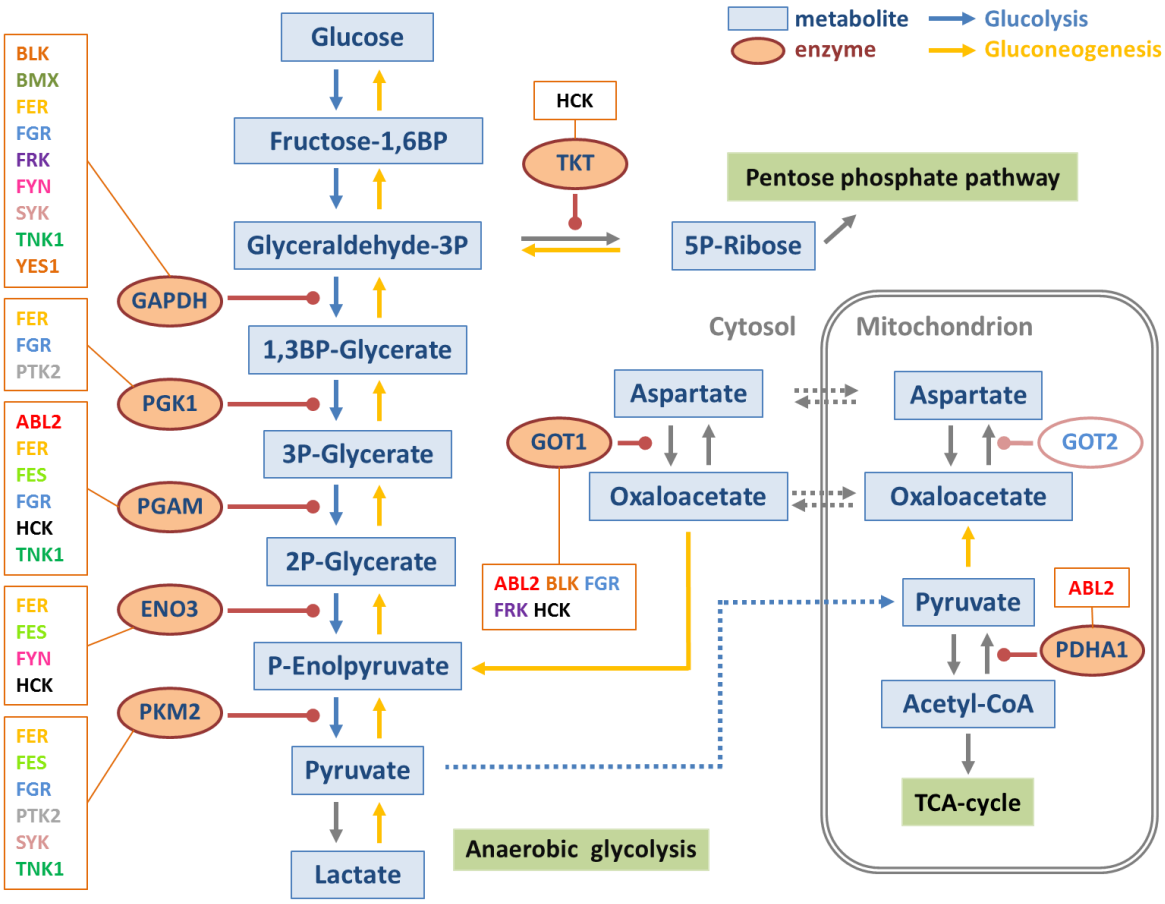


Figure 50: Reduced schematic representation of glucose metabolism. Metabolites are boxed in blue, metabolic enzymes targeted by NRTKs in red ovals. For each metabolic enzyme, an orange box lists the homology predicted kinases. Blue arrows indicate (reversible) steps in glycolysis, yellow arrows steps in gluconeogenesis, dashed arrows indicate translocations.

4.8.1.3. NRTKs are targeting enzymes linked to the glycolytic pathway

Besides glycolytic enzymes, TLK1, the yeast ortholog of TKT, an enzyme linking the glycolytic pathway to the non-oxidative pentose phosphate pathway, was exclusively targeted by HCK on Y294 corresponding to Y275 in human TKT. Interestingly, human TKT was additionally predicted to be modified at Y275 by ABL2 and BLK via sequence motif (Table 7). Metabolic control analysis by Cascante et al. (2000) suggested that TKT expression is highly up-regulated in cancer cells, tumor growth correlates with TKT activity, and application of TKT inhibitors resulted in a decrease of tumor growth. Previously identified as a key factor in amino acid controlled purine synthesis, Saha et al. (2014) could show that TKT is post-transcriptionally regulated by an activating phosphorylation of T382 by AKT S/T kinase in amino acid deprived HeLa cells using various biochemical assays and

phospho-proteomics. Thus, TKT targeting by HCK, ABL2, and BLK may modify its enzymatic activity as well. Furthermore, cytosolic aspartate aminotransferase (AAT2 in yeast, GOT1 in human), but not the mitochondrial version AAT1 (GOT2 in human) was targeted on Y70 (Y71 in human) by ABL2, BLK, FGR, FRK, and HCK. GOT1 and GOT2 convert glutamine-derived aspartate to oxaloacetate which can be processed first to malate and further to 2P-enolpyruvate and thus feeds into the gluconeogenesis and glycolysis pathways. Son et al. (2013) could show that GOT1 knock-out in pancreatic cancer cells, but not in healthy cells, resulted in reduced proliferation. Metabolomics analysis suggested that KRAS-induced tumor growth is regulated by alternative glutamine metabolism during GOT1-dependent maintenance of the cancer cell's redox state. Thus, targeting of GOT1 by SFKs and ABL may elevate GOT1 enzyme activity and hence results in cell proliferation by increased glutamine metabolism. Interestingly, yeast PDA1, the E1 alpha subunit of the pyruvate dehydrogenase complex (PDC) converting pyruvate to acetyl-CoA before entering the citric-acid cycle, was targeted exclusively by ABL2 on Y321 corresponding to Y301 in human PDHA1. Fan et al. (2014) observed PDHA1 phosphorylation on Y301 by mutational analysis and immuno-blotting in growth factor stimulated human cancer cells and tissues, but not in normal cells. The authors proposed that the tyrosine modification inhibits PDHA1 by blocking substrate binding. The identification of Y301-phosphorylated PDHA1 in BCR-ABL positive leukemic cells prompted the authors to test whether ABL can modify this site in an *in vitro* kinase assay. They could show that the ABL indeed phosphorylates Y301 which is in agreement with the measured targeting of the yeast orthologous site by ABL2 and the resulting homology based prediction for human PDHA1. Thus, the findings of Fan et al. (2014) further indicate that tyrosine phosphorylation KSRs man be inferred directly via NRTK targeting of proteins conserved between yeast and human and that these modifications may have functional consequences in driving proliferative cells into aerobic glycolysis.

4.8.1.4. NRTKs are targeting glycolytic enzymes

4.8.1.4.1. GAPDH phosphorylation

In the following paragraphs literature evidence in support of NRTK targeting of glycolytic enzymes is discussed in more detail. Of all glycolytic enzymes, GAPDH was targeted most intensively on four tyrosine by a total of nine assayed NRTKs. Indeed, apart from FAK (PTK2), members of all NRTKs families were targeting GAPDH whereat TNK1 was the only NRTK targeting all four targeted, conserved yeast GAPDH pY-sites. SYK targeted three pY-sites, but not Y41. Y41 was modified by FER which showed mutual exclusivity to SYK in overall glycolytic enzyme targeting on a site level and the SFKs FGR, FYN, and YES1 (Figure 51).

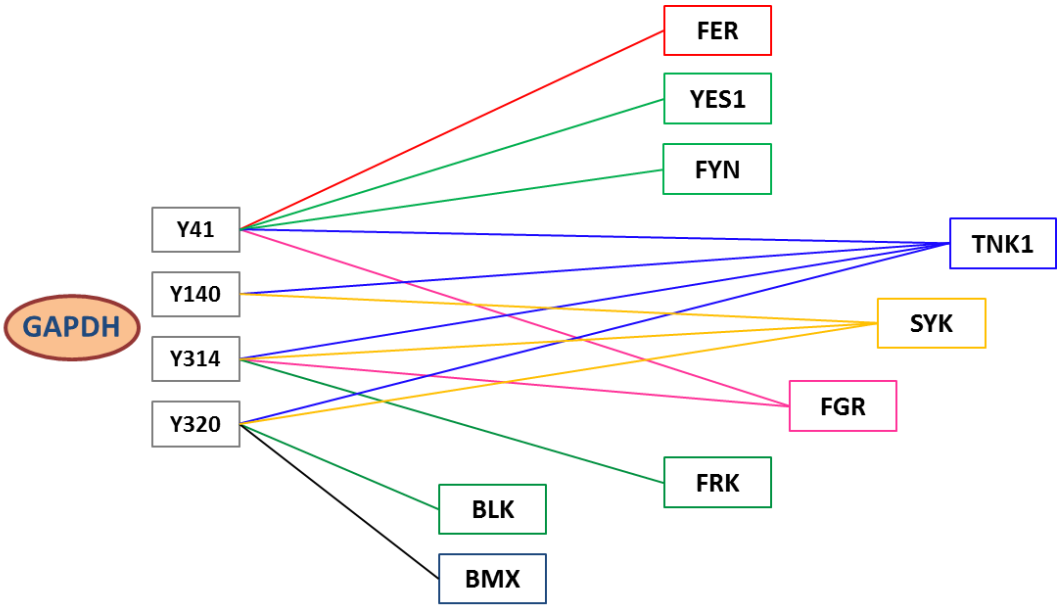


Figure 51: Scheme showing the overlap of inferred NRTK targeting of GAPDH tyrosine residues Y41, Y140, Y314, and Y320.

This pattern is in agreement with the reported moonlighting function of the ubiquitous GAPDH. It is well established that GAPDH has many other functions outside glycolysis likely regulated by PTMs as reviewed by Tristan et al. (2011). Therefore, it is likely that some of the pY-sites observed in yeast and predicted for human may have functional consequences not related to glycolysis at all. One of the many functions of GAPDH is the regulation of vesicle transport from the endoplasmatic reticulum to the Golgi apparatus (Tisdale and Artalejo, 2006). GAPDH is forming a complex on pre-Golgi intermediates termed vesicular tubular clusters with small GTPase Rab2, atypical protein kinase C and SRC in a tyrosine phosphorylation dependent manner (Tisdale and Artalejo, 2006). Tisdale and Artalejo (2007) conducted mutational studies and could show that phenylalanine substitution of Y41 abolished SRC-dependent phosphorylation of GAPDH and blocked the formation of the complex with Rab2 and atypical PKC resulting in impaired vesicle formation and transport in the early secretory pathway. Even though SRC did not target GAPDH in yeast, other NRTKs of different NRTK families may also modify this critical tyrosine for signaling complex assembly. Functions fulfilled by the other three GAPDH pY-sites, targeted by subsets of certain NRTKs in yeast as shown in Figure 51, have not been reported, yet.

In the two subsequent glycolytic steps, two of the homology predicted KSRs, PGK1 Y76 targeted by FGR and PGAM1 Y92 targeted by ABL2, were experimentally verified in this thesis in a developed *in vitro* kinase assay using yeast and a mass spectrometry read-out. Both NRTK substrates are key enzymes in the glycolytic pathway, share high amino acid sequence similarity (76.6% for PGK1 and

64.6% for PGAM1) between yeast and human and expression of the human version in yeast was sufficiently strong for enrichment and measurement by the established validation assay.

4.8.1.4.2. PGK1 phosphorylation

PGK1 is a glycolytic enzyme that catalyzes the conversion of 1,3-diphosphoglycerate to 3-phosphoglycerate and its ortholog in yeast was targeted on two sites Y76 and Y196 by FER and FGR whereas Y76 additionally by PTK2. Similar to other glycolytic enzymes such as PGAM1, PGK1 showed increased expression in solid tumors for the production of ATP during anaerobic conditions such as hypoxia and promotes metastasis as reported by Wang et al. (2007) and Ahmad et al. (2013), for instance. Urokinase-type plasminogen activator receptor (uPAR) expression is induced by the receptor's own antigen the serine protease urokinase-type plasminogen activator (uPA). uPAR expression results in plasminogen activation in the extra-cellular matrix which is a critical event in cell invasion and tumor metastasis under hypoxic conditions (Sullivan and Graham, 2007). Shetty et al. (2010) discovered PGK1 binding to uPAR mRNA which led to uPAR mRNA stabilization and hence cell surface expression of uPAR. The authors could show by mutational analysis and treatment of cells with uPA that the N-terminus of PGK1, and not the C-terminus providing its glycolytic activity, is required for binding uPAR mRNA and that this mechanism is dependent on the phosphorylation of Y76 of PGK1, but not Y196. FGR was predicted to target Y76 and Y196 and by the established mass spectrometry coupled *in vitro* kinase assay it was experimentally verified that FGR can phosphorylate Y76, but interestingly, not Y196 (Figure 35). Thus, it can be hypothesized that FGR together with FER and FAK (PTK2) may play a role in uPA induced uPAR-dependent plasminogen activation resulting in cancer progression by a glycolysis-independent mechanism. This does not exclude the possibility of NRTK targeting of PGK1 may change its glycolytic activity as well.

4.8.1.4.3. PGAM phosphorylation

PGAM catalyzes the reversible reaction of 3-phosphoglycerate (3-PGA) to 2-phosphoglycerate (2-PGA), another important step in the glycolytic pathway. PGAM is expressed as isoform 1 in brain tissue and isoform 2 in muscle tissue and both isoforms are reported to be phosphorylated in human on Y92 (phosphositeplus.org). In yeast, ABL2, FER, FES, FGR, HCK, and TNK1 targeted the PGAM1 ortholog glycerate phosphomutase 1 (GPM1) on the corresponding residue Y90. Co-expression of ABL2 and PGAM1 in yeast and subsequent enrichment identified phosphorylation of Y92 of human PGAM1 by MS (Figure 34) and only ABL2 can be the responsible enzyme due to the absence of NRTKs in yeast. The positive result by testing ABL2 may suggest that Y92 of both isoforms of PGAM can be targeted by ABL2 and also by the other NRTKs predicted by homology which are members of three different NRTK families – FES, SRC, and ACK. Hitosugi et al. (2012) reported up-regulated expression

of PGAM1 in several types of cancers, including leukemia, and it was suggested this was in part due to loss of TP53 activity which is a negative transcriptional regulator of PGAM1. Additionally, the researchers could show that in cancer cells not only PGAM1 levels, but also PGAM1 activity is increased resulting in elevated glycolysis and hence may contribute to "Warburg effect". Later, Hitosugi et al. (2013) performed mutational, crystallographic, and MS-based characterization of PGAM1 phosphorylation and suggested that Y92 is intrinsically required for PGAM1 activity as phenylalanine substitution of Y92 rendered PGAM1 inactive. Proximal to the rare case of a phosphorylated histidine (H11) which is also essential for PGAM1 activity, Y92 is located at the active site and may be required for cofactor and substrate binding similar to H11. The authors further could show that phosphorylation of PGAM1 Y26 which is as well in spatial proximity to the catalytic site resulted in increased PGAM1 enzyme activity and promoted cancer cell proliferation and tumor growth. Thus, Y92 phosphorylation may influence PGAM1's catalytic activity and may be also important in cancer progression. In BCR-ABL positive leukemic cells ABL kinase activity is increased and deregulated by fusion to breakpoint cluster region (BCR) via chromosomal translocation (Heisterkamp et al., 1990). Therefore, the observed targeting of PGAM1 by ABL2 may be an important factor in proliferation in this type of cancer cells.

4.8.1.4.4. ENO and LDHA phosphorylation

The last steps in glycolysis involve the metabolic enzymes enolase (ENO), pyruvate kinase, and lactate dehydrogenase A (LDHA). Already in 1984, Cooper et al. (1984) performed kinase assays employing radiolabeling and 2D-gel electrophoresis and reported tyrosine phosphorylation of PGAM and to a lesser extent of ENO by FER, FES, ABL, SRC, and YES1 *in vitro*. Additionally, modification of a single tyrosine of ENO and LDHA *in vivo* by FER and FES was shown. Cooper and co-workers suggested a region homologous to the sequence surrounding histidine 43 in yeast enolase which does not agree with the measured phosphorylation of Y131 of yeast ENO1 (corresponding to Y131 in human ENO3) by FER, FES, FYN, and HCK as the residues are too distant. FYN and HCK were not tested to modify ENO by Cooper et al. (1984), but showed great targeting specificity overlap in the yeast measurements to SRC and YES1. Additionally, Y131 in human ENO3 was predicted to be modified by ABL2 and LYN matching the sequence motif generated in the yeast experiments (Table 7).

4.8.1.4.5. PKM phosphorylation

The final step of glycolysis is catalyzed by pyruvate kinase converting phosphoenolpyruvate to pyruvate which is subsequently converted by LDHA to lactate in anaerobic glycolysis. There are four isoforms of pyruvate kinase (M1, M2, L, and R) which are expressed in different tissues with differing levels of activity. PKM1 is the predominant isoform expressed in tissues with high energy

consumption such as brain and muscle whereas PKM2 is rather expressed in the lung and in proliferating cells which are synthesizing vast amounts of nucleic acids such as embryonic cells or tumor cells (Mazurek, 2011). As reviewed by Mazurek (2011), pyruvate kinase isoforms form a tetramer quaternary structure whereas PKM2 may occur in both tetrameric and dimeric form. Kinetic studies on PKM2 determined that the tetramer provides high-affinity and the dimer low affinity for its substrate phosphoenolpyruvate. Therefore, together with spatial proximity to other glycolytic enzymes in the GEC, presence of the highly active tetrameric form results in glucose degradation primarily to pyruvate and lactate as observed by Otto Warburg (Mazurek, 2011). It was previously assumed this metabolic reconfiguration results from a switch in the expression of PKM1 to PKM2 and suggest to be an important factor for malignant transformation (Mazurek et al., 2005, Christofk et al., 2008a). This hypothesis was rejected by Bluemlein et al. (2011) who performed absolute quantification of pyruvate kinase isoforms by a method employing isotope labeling and mass spectrometry in 25 human malignant cancers, 18 tissue-matched controls, 12 cancer cell lines, 4 non-cancer cell lines, and 6 benign oncocytomas. The comparison of pyruvate kinase isoforms levels between malignant and normal cells demonstrated that predominance of PKM2 is not a result of a switch in isoform expression. The authors suggested that it may be PKM2 oligomer formation controlled by PTMs which is driving cells into lactate production by changing PKM2 activity as proposed by the findings of Hitosugi et al. (2009) which are described in detail below. It appears as a paradox that in tumor cells producing vast amounts of lactate, the prevalent form of PKM2 is the nearly inactive dimer. This can be explained by the benefits of tumors using glycolytic intermediates for anabolic biosynthesis while increased lactate production and excretion is an effective way to recycle nicotinamidadenindinukleotid (NAD+) needed by GAPDH and for the excretion of excess hydrogen (Mazurek et al., 2005). A central role of PKM2 in cancer metabolism was also proposed by Christofk et al. (2008a) who knocked-down PKM2 in several human cancer cell lines using short hairpin RNA and replacing it with PKM1. Replacement resulted in the reversal of the Warburg effect as observed by reduced lactate production, increased oxygen consumption, and diminished tumor growth in a mouse xenograft. The researchers identified PKM2 as a phospho-peptide binding protein by conducting proteomic screening using HeLa cells and SILAC employing a phosphorylated and un-phosphorylated peptide library (Christofk et al., 2008b). Additionally, Christofk et al. (2008b) solved a crystal structure of PKM2 and could show that Fructose-1,6BP, a positive allosteric regulator of PKM2, is released upon binding of tyrosine phosphorylated peptides derived from ENO and LDHA, and a peptide matching the SRC sequence motif. The release of Fructose-1,6BP keeps PKM2 in its inactive, dimeric form by preventing tetramer formation (Christofk et al., 2008b). Later, Hitosugi et al. (2009) conducted MS-based phospho-proteomics using murine hematopoietic cells stably expressing the constitutively active fusion tyrosine kinase ZNF198-FGFR1 and immuno-affinity

enrichment employing an anti-phospho-tyrosine antibody (the same commercially available protocol adapted for the yeast experiments here) and determined Y83, Y105, Y148, Y175, Y370, and Y390 to be phosphorylated. Subsequent co-expression of a constitutively active mutant form of ZNF198-FGFR1 and PKM2 tyrosine to phenylalanine point mutants in 293T cells demonstrated that the Y105F PKM2 mutant had significantly increased PKM2 activity compared to wild-type in a FGFR1-dependent manner. Of note, PKM2 Y83F was the only mutant that slightly decreased PKM2 activity, but non-significantly, compared to wild-type. Similar to Christofk et al. (2008b), Hitosugi et al. (2009) could measure release of Fructose-1,6BP from PKM2 resulting in its inactivation by addition of a phospho-peptide based on the PKM2 sequence surrounding Y105 of PKM2. Thus, it was suggested that a Y105 phosphorylated PKM2 molecule may disrupt tetramer formation by interaction with other, non-phosphorylated PKM2 molecules. Furthermore, they found Y105 to be specifically modified in 10 of 12 cancer cell lines and that expression of the PKM2 Y105F mutant in cancer cells led to decreased proliferation under hypoxic conditions, increased oxidative phosphorylation, and decreased lactate production under normoxic conditions. Additionally, Hitosugi et al. (2009) observed reduced tumor growth in mouse xenografts as compared to less active wild-type PKM2. Altogether, their findings suggest a pivotal role of PKM2 Y105 phosphorylation in the Warburg effect. Interestingly, PKM was targeted in the yeast experiments on two tyrosine by two distinct sets of human NRTKs - Y83 by FER, FES, FGR, and TNK1 and Y466 by FAK (PTK2) and SYK. Both modification sites were reported in the phosphositeplus.org database for PKM1 and PKM2 as peptide sequences are identical. One may speculate that Y83 phosphorylation is also required for PKM2 oligomer formation as mutation results in a slight loss of activity as mentioned above. Phosphorylation of Y466 by SYK and FAK (PTK2) may also change PKM2 conformation or oligomer formation, respectively, but could together with Y83 also be important in the assembly of the GEC.

In summary, human KSRs were predicted via phosphorylation site conservation between yeast and human. All glycolytic enzymes were targeted in yeast and orthologous sites were reported to be phosphorylated in human. The impact of glycolytic enzymes in driving aerobic glycolysis and their regulation by PTMs was briefly discussed. Overlapping and specific modifications of glycolytic enzymes by cytoplasmic PTKs were measured in yeast. Subsequently, the findings were transferred to human proteins - in part experimentally verified using a kinase assay and backed by literature. Extensive modification of all glycolytic enzymes by the several cytoplasmic PTKs presented here suggests that (aerobic) glycolysis and oncogenic cytoplasmic tyrosine kinase signaling may be more highly inter-linked than previously assumed

4.8.2. Primary sequence specificities obtained from yeast can be used to predict human substrates

Analysis of the amino acid sequence surrounding the pY-sites within each NRTK target set in yeast enabled generation of novel linear amino acid sequence motifs for each tested NRTK. Unlike previously published motif-based human kinase-substrate predictors, the motifs were generated from folded, full-length yeast substrates naturally containing 3D structural information given by the cellular environment of the alive lower eukaryote. Thus, the motifs may represent distinct NRTK substrate specificities when compared to previously published versions. Analysis of reported human pY-sites using optimized motifs yielded 117 predictions on average for each NRTK whereas up to 80 percent may be true positives (Section 3.10). The strength of assigning NRTKs to human pY-sites using a comprehensive set of NRTK motifs generated in the same experimental set-up will be demonstrated on a few exemplary findings in the following paragraphs.

4.8.2.1. Motif predictions indicate NRKT targeting of cell cycle regulators

Over 1000 measured human phosphorylation sites were assigned to a single NRTK whereas overlapping predictions may enable to infer NRTK co-operation in cellular processes within or in-between NRTK families or suggest non-overlapping targeting specificity due to differential tissue expression. For instance, cyclin-dependent kinases 16 and 17 (CDK16 and CDK17) were the two predicted human targets with the highest number of kinases predicted to modify a single site.

Symbol	GeneID	RefSeqID	Site	Sequence	Motif predicted kinases
CDK1	983	NP_001777	Y15	EKIGEGT**Y**GVVYKGR	BLK, FER, FGR, FYN, HCK
CDK2	1017	NP_001789	Y15	EKIGEGT**Y**GVVYKAR	BLK, FER, FYN
CDK3	1018	NP_001249	Y15	EKIGEGT**Y**GVVYKAK	BLK, FER, FYN
CDK5	1020	NP_004926	Y15	EKIGEGT**Y**GTVFKAK	BLK, FYN, SRC
CDK11A	728642	NP_076916	Y434	NRIEEGT**Y**GVVYRAK	BLK, FYN, LYN
CDK11B	984	NP_277021	Y436	NRIEEGT**Y**GVVYRAK	BLK, FYN, LYN
CDK13	8621	NP_003709	Y362	LPRSPSP**Y**SRRRSPS	HCK
CDK13	8621	NP_003709	Y716	GIIGEGT**Y**GQVYKAR	BLK
CDK14	5218	NP_036527	Y128	EKLGEGS**Y**ATVYKGK	FER, SRC
CDK15	65061	NP_631897	Y63	EKLGEGS**Y**ATVYKGI	FER
CDK16	5127	NP_148978	Y182	DKLGEGT**Y**ATVYKGK	BLK, FER, FGR, FYN, HCK, LYN, SRC, YES1
CDK17	5128	NP_002586	Y203	EKLGEGT**Y**ATVYKGR	BLK, FER, FGR, *FRK*, FYN, HCK, SRC, YES1
CDK18	5129	NP_997668	Y185	DKLGEGT**Y**ATVFKGR	BLK, FGR, FYN, HCK, SRC
CDKL5	6792	NP_003150	Y171	NNANYTE**Y**VATRWYR	FGR, *FRK*
CDKN1B	1027	NP_004055	Y88	KGSLPEF**Y**YRPPRPP	FER, *FRK*

Table 11: Motif predicted human CDK pY-site targeting by SFKs, FER (underline) and FRK (*italic*).

Cyclin-dependent kinases (CDKs) belong to the family of S/T kinases that mainly regulate cell cycle progression. Interestingly, the observation that Y182 of CDK16 matched the sequence motif of all tested SRC-family kinases (SFKs) may suggest a crucial signaling event in cell cycle control involving an entire NRTK family. Indeed, a total of 15 CDKs or CDK modulators were predicted to be targeted by SFKs, SFK-related kinase FRK, and FER as outlined in Table 11. None of the sequences appears to be the kinase activation loop. Apart from SFKs, the only other two NRTKs which are motif predicted to modify CDKs were FRK and FER which also showed strong targeting specificity overlap in yeast (Figure 15). To date, FER and FRK are only known to indirectly regulate the cycle by phosphorylation of CDK effectors (Pasder et al., 2006, Hua et al., 2014). Of note, phosphorylation of Y19 of CDK3 was also predicted by site conservation between yeast and human. This residue was targeted in yeast by the SFKs BLK, FGR, FYN, HCK, SRC, YES1, and additionally by ABL2 and TNK1. Furthermore, Y19 of CDK3 is conserved in CDK1 and CDK2 however, phenylalanine substituted in CDK5. In this regard, CDK5 was suggested to be different from all other CDKs having a function in various neuronal activities unrelated to the cell cycle (Kobayashi et al., 2014). On the other hand, CDK1, CDK2, CDK3, and CDK5 all have the residue Y15 motif predicted to be a FYN substrate which was reported earlier: Sasaki et al. (2002) used a CDK5 Y15-recognizing antibody to detect Y15 phosphorylation in HEK293T lysate upon co-expression of GST-tagged constructs of mutated and constitutively active mouse Fyn and mouse CDK5. Sasaki and colleagues proposed that phosphorylation of Y15 of CDK5 has an activating effect and may play a role in neurite and spine retraction, dendrite outgrowth, and neuron death. This is in contrast to the proposal of an inhibitory effect of Y15 phosphorylation of CDK1, CDK2, and CDK3. In regard of CDK1, which forms a complex with cyclin B, phosphorylation of Y15 retains the complex in an inactive state until it becomes dephosphorylated at the onset of cell cycle M-phase (Kobayashi et al., 2014), for example. Although Y15 surrounding sequences are very similar between homologs, some differences in NRTK motif predictions were observed, the *in vivo* relevance of which remains elusive. Y15 of CDK1, CDK2, and CDK3 was predicted to be modified by BLK, FER, and FYN whereas FGR and HCK were additionally predicted to modify Y15 of CDK1 at the chosen cut-off. Y15 of CDK5 was predicted to be targeted by BLK, FYN, and SRC however, not FER. Previous observations that ABL phosphorylation of CDK5 resulted in p35 (cyclin-dependent kinase 5, regulatory subunit 1; CDK5R1) binding and hence in the activation of CDK5, prompted Kobayashi et al. (2014) to investigate whether other NRTKs could activate CDK5 in a similar way. Therefore, they ectopically expressed constitutively active FYN and SRC in COS-7 cells and generated a CDK5 phospho-Y15 specific antibody which is not cross-reacting with COS-7 cell expressed monkey CDK1 and CDK2 phospho-Y15. Thereby, they could show that both NRTKs can modify CDK Y15 *in vivo*. Thus, two of three motif-based kinase-substrate predictions made here concerning Y15 of CDK5 are supported by literature evidence. As a next step, Kobayashi et al. (2014) co-expressed CDK5 binding

proteins including p35 and observed inhibition of Y15 modification by FYN upon binding of CDK5 regulators which was independent of phosphatase activity. The researchers could further show that overexpression of FYN results in activation of CDK5, however, not due to Y15 phosphorylation but due to higher p35 levels by reduced p35 degradation. Although it was determined that substitution of Y15 impaired CDK5 activity in cultured neurons, phosphorylation of Y15 was not influencing CDK5 activity which may be due to NRTK-dependent up-regulation of CDK5 activating binders such as p35. Another regulator of CDKs is CDKN1B (also known as p27 or KIP1) which inhibits cyclin/CDK complexes upon binding by remodeling the CDK catalytic cleft and blocking the ATP binding pocket. Grimmler et al. (2007) reported that Y88 residing in the CDK binding domain of CDKN1B is phosphorylated by SFKs SRC and LYN, and the BCR-ABL oncoprotein. Phosphorylation of Y88 caused re-activation of cyclin/CDK complexes by ejecting CDKN1B from the catalytic cleft and ATP binding site whereas CDKN1B remained bound to the complex. The authors further demonstrated that the conformational change in CDKN1B caused by Y88 phosphorylation transformed the CDK inhibitor to a CDK substrate. This led to CDKN1B degradation during the normal cell cycle and hence Y88 targeting by activated tyrosine kinases in cancer may result in proliferation by CDK activation – constituting a direct link between cell cycle and oncogenic tyrosine kinase signaling. Interestingly, the only two NRTKs motif predicted to modify Y88 of CDKN1B were FER and FRK, but not ABL2, SRC, or LYN suggesting an involvement of FER and FRK in CDK regulation by putative CDKN1B phosphorylation.

4.8.2.2. NRTKs are motif predicted to be substrates themselves indicating inter-NRTK regulation complexity

Among the top motif scoring human pY-sites many NRTKs were identified as NRTK substrate themselves. An overview of inter-kinase regulation by prediction of motif matching sites o NRTKs is given in Figure 52 demonstrating its complexity. Apart from FER and TNK1, each tested NRTK is predicted to modify or to be modified by at least one other NRTKs. Notably, for FRK a total of six tyrosine modifications by five SFKs were predicted whereas FRK was the only NRTK predicted to target two sites on a single kinase (YES1) and FRK itself was not predicted to be targeted by other NRTKs. Similarly, both ABL2 and BLK were predicted to phosphorylate three other NRTKs on a single site and both NRTKs were not motif predicted to be NRTK substrates. YES1 and SYK were the NRTK substrates with the highest number of three predicted modification sites. Interestingly, YES1 was not predicted to target NRTKs and SYK was exclusively predicted to target Y419 of PTK2. SYK appears to be relatively higher regulated by other NRTKs as the highest number of five predicted NRTKs may modify SYK. Of note, phosphorylation of Y630 and Y631 at the C-terminal tail of SYK regulates SYK activity and function and were predicted to be targeted by FES, BMX, and SRMS and not by known regulators including SFKs or SYK (de Castro et al., 2010). Interestingly, no classical activating and

inhibiting pY-sites such as the Y416 auto-phosphorylation site and the Y527 CSK regulatory site for SFKs (Okada, 2012) were involved. NRTK auto-phosphorylation sites were determined in the yeast experiments by querying the mass spectrometry output for modified human sequences and hence lead to the discovery of novel NRTK sequence modifications which may regulate NRTK activity as shown for SFKs (Hinkle et al., 2015). There was not a single pair of motif predictions to indicate that two kinases mutually phosphorylate. Inter-kinase regulation by phosphorylation can be hypothesized to be even more complex than presented here as only half of all NRTKs were analyzed and inter-NRTK targeting is likely to be also regulated motif-independent by contextual information such as adaptor proteins. In summary, motif prediction based on yeast data not only enabled predictions specific for NRTKs and NRTK family substrates, but also may be used to decipher inter-kinase relationships.

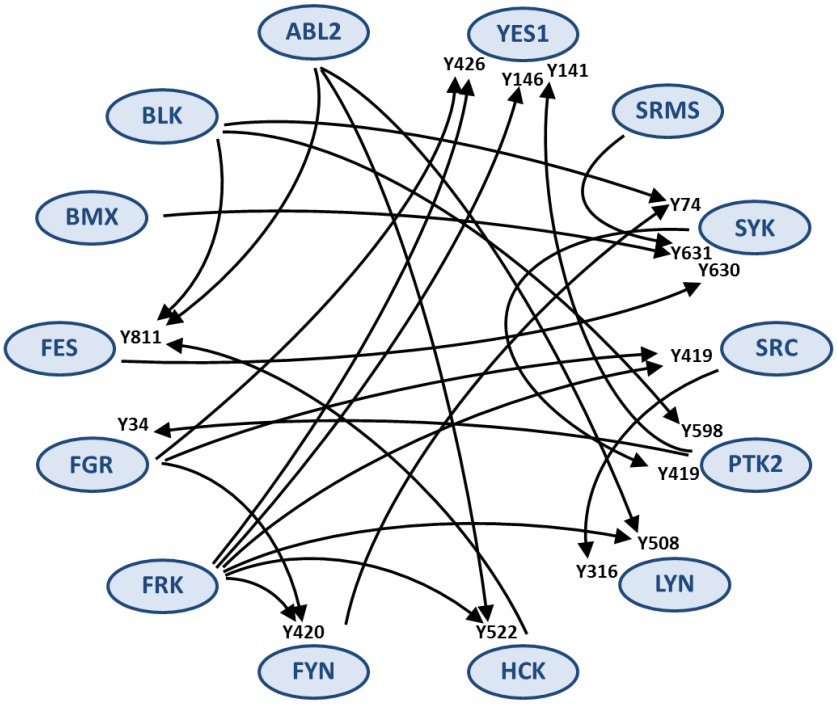


Figure 52: NRTKs as NRTK substrates predicted by motif scoring of human pY-sites.

4.8.2.3. Motif predicted EIF2S1 modification by FGR is validated

The only motif predicted human substrate tested that showed sufficient expression amenable to targeted enrichment by the established kinase assay was eukaryotic initiation factor 2 subunit 1 (EIF2S1). The primary sequence spanning Y150 in EIF2S1 was predicted to match the motifs of three SFKs, BLK, FGR, and FYN, with high probability. Surprisingly, co-expression of EIF2S1 with FGR resulted in phosphorylation of both Y150 and Y147 of EIF2S1 by FGR. Thus, the motif prediction of FGR for targeting EIF2S1 Y150 was experimentally validated and in addition, FGR could be assigned to

modify Y147 which may be phosphorylated sequentially by FGR due to close proximity to the motif matching site. The consequences of EIF2S1 phosphorylation by SFKs are unknown.

4.8.2.4. Motif predictions recapitulate kinase dependencies in molecular mechanisms and signaling as shown for CRK and related molecules

Another motif predicted target of ABL2 was v-crk avian sarcoma virus CT10 oncogene homolog (CRK). CRK is an important cytoskeleton regulator and a highly investigated molecule in modular kinase signaling. CRK contains non-catalytic SH2 and SH3 domains similar to the modular domains of SRC-, FRK-, CSK-, ABL-, and TEC-family cytoplasmic PTKs however, is lacking intrinsic tyrosine kinase activity. CRK was shown to regulate tyrosine kinase activity in *trans* as opposed to intra-molecular *cis*-regulation of NRTKs via their own SH2 and SH3 domains (Kumar et al., 2014). As reviewed by Kumar et al. (2014), there are three cellular homologs, two splice variants CRKI and CRKII, and the structurally related, but functionally different CRK-like (CRKL). All CRK proteins contain a phospho-tyrosine binding SH2 domain followed by a poly-proline type 2 motif (PxxPxK/R) binding N-terminal SH3 domain (SH3N) and an atypical C-terminal SH3 domain (SH3C) which does not bind poly-proline motifs. The SH3N domain was shown to bind specific guanine nucleotide exchange factors (GEF) resulting in activation of GTPases and the SH2 domain was determined to bind specific tyrosine phosphorylated proteins such as p130/Cas and paxillin (PXN) which are signaling molecules important for cell migration (Kumar et al., 2014). It was suggested that the interaction with CRK SH2 stabilizes the phosphorylation status of its binding partners or even resemble a molecular “sensor” for tyrosine phosphorylation levels in cells (Kumar et al., 2014). CRK and CRKL were found to be dysregulated in human malignant tumors and CRK abundance correlates with disease outcome and many disease relevant proteins were identified binding to its SH2 and SH3N domains (Kumar et al., 2014). CRK conformation and signaling is dynamically controlled by tyrosine phosphorylation as reported already over 20 years ago by Feller et al. (1994). The authors performed *in vitro* kinase assays using truncation and point-mutants of CRK showing that phosphorylation of Y221 by ABL kinases auto-inhibits CRK by formation of an intra-molecular clamp. Later, Sriram et al. (2011) conducted kinase assays, pull downs, and high-throughput SH2 profiling approaches and hence demonstrated that CRK can be *in vitro* and *in vivo* phosphorylated by ABL at Y251 in a conserved RT loop structure of the SH3C domain. Y251 phosphorylation resulted in ABL *trans*-activation of ABL1 isoform B by binding to phosphorylated Y251 with its SH2 domain. Recently, Sriram et al. (2014) could show via mass spectrometry measurements and by generating affinity-purified phospho-tyrosine-specific antibodies that CRK, when incubated *in vitro* with recombinant ABL, was phosphorylated at Y221 the SH3N-SH3C linker region, at Y239 at the SH3C boundary and at Y251 in the RT-loop. Different stimuli such as EGF and fibronectin, or over-expression of growth factor

receptors (receptor protein tyrosine kinases) led to phosphorylation of one of the three tyrosine residues or specific pair-wise combinations thereof. Furthermore, the authors performed isothermal calorimetric (ITC), pull-downs, and NMR experiments and thus were able to show that auto-inhibition of CRK by SH2 domain binding upon phosphorylation of Y221 did not necessarily render SH3N in-accessible. Moreover, it was demonstrated that CRK activity was not abolished when Y239 and/or Y251 were concomitantly phosphorylated. Phosphorylation of Y239 and Y251 resulted in a shift from CRK SH2-SH3N domain utilization to a non-canonical phosphoSH3C-SH3N signaling mode whereby phosphoSH3C formed a *de novo* binding interface for SH2 domains of ABL1/ABL2 and c-SRC tyrosine kinase (CSK). Using peptides harboring phospho-Y239 and phospho-Y251 Sriram and coworkers screened a SH2 domain library and observed ABL1/ABL2 and SCK/SHCB SH2 domain binding to both tyrosine residues and also a preference for Y239 by the SH2 domain of CSK. CSK is a negative regulator of SFKs and co-transfection of SFKs with CRK in HEK 293T cells confirmed phosphorylation of Y239 by activated HCK, SRC, LYN, and BLK. The model of CRK regulation by domain conformation as proposed by Sriram et al. (2014) and extended by NRTK motif predictions is shown in Figure 53 and explained below.

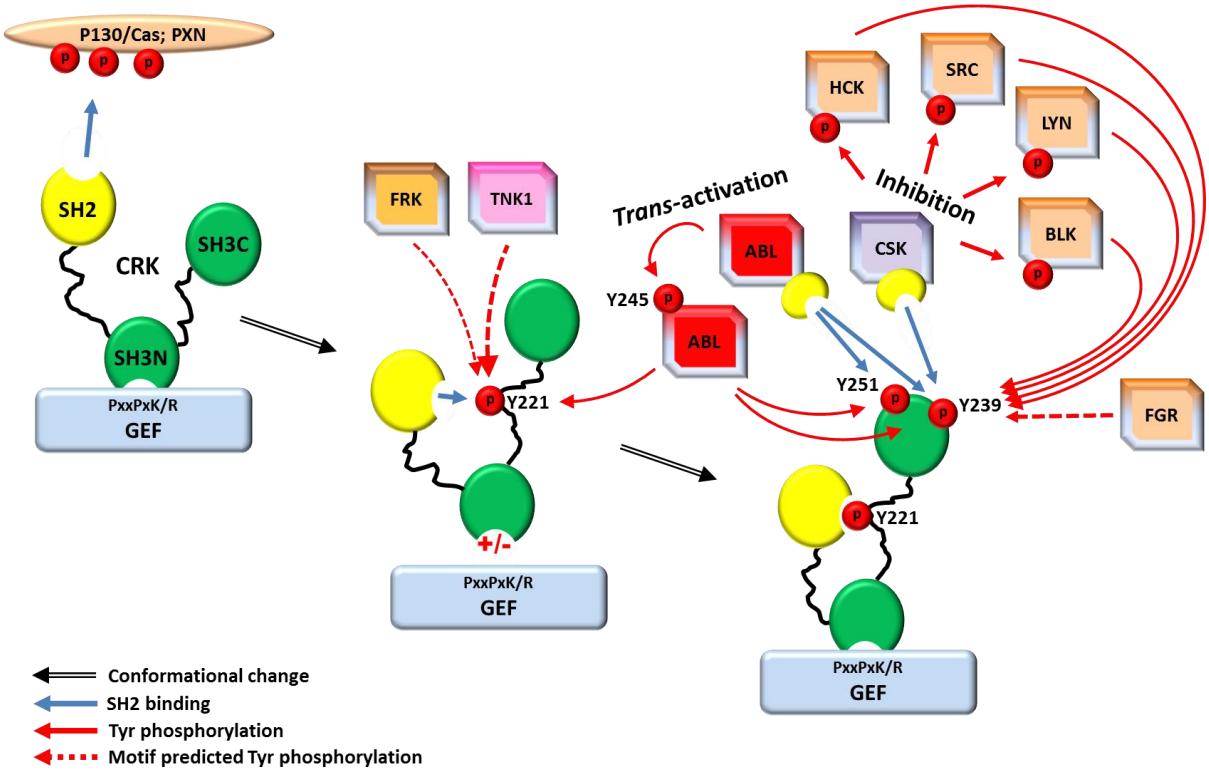


Figure 53: Model of CRK modular signaling including a non-canonical signaling mode by phosphorylation of Y239 and Y251 on SH3C modified from Sriram et al. (2014). Non-phosphorylated CRK binds to signaling molecules via its SH2 domain and its SH3N domain (left). Y221 phosphorylation results in CRK auto-inhibition by conformational change (center). SH2 binding can inhibit or retain SH3N binding to GEFs. Iterative phosphorylation of Y221 and Y239/Y251 creates additional, novel binding surfaces for NRTK SH2 domains resulting in ABL transactivation and Src-family kinase inhibition (right). Motif predicted modifications are indicated by dashed arrows. ABL2 phosphorylation of Y221 and Y239 is shown by continuous arrows as it was motif predicted and observed previously by others.

While CRK was strongly expressed in the yeast kinase assay system, upon tryptic digest ABL2 motif predicted sites Y221 and Y239 reside on a very long peptide (YRPASASVSALIGGNQEGSHPQPLGGPEPGPY(ph)AQPSVNTPLPNLQNGPIY(ph)AR) which was not observable in the performed mass spectrometric measurements. Nevertheless, reported investigations on CRK modification, in particular the work of Sriram et al. (2014), validated the motif based predictions for ABL2 activity on Y221 and Y239. The only other kinase which was motif predicted to modify CRK Y221, but not Y239, with a motif score above cut-off was TNK1. Strikingly, TNK1 was the only kinase predicted to phosphorylate the corresponding inhibitory site Y207 on CRKL. Y207 of CRKL was shown previously to be modified by ABL1 *in vitro* by co-transfecting in HEK 293T cells with activated ABL1 with CRKL wild-type and phenylalanine mutants (Zipfel et al., 2004). TNK1 only has a single SH3 domain and no SH2 domain (Figure 7). Notably, two additional NRTKs may also phosphorylate CRK having a motif score (accuracy value) assigned which is close to however, below the chosen accuracy cut-off 0.995. The sequence around Y221 may fit the motif of FRK (ACC 0.9930). Y239 may also be targeted by FGR having a motif score just below cut-off (ACC 0.9950) which would agree with the findings of Sriram et al. (2014) who demonstrated that Y239 is a target of SFKs. Both FGR and FRK have a single SH2 domain and a single SH3 domain N-terminal to the kinase domain (Figure 7) which may bind CRK.

CRK phosphorylation was reported to be strongly increased upon ABL2 activation by another molecular adapter molecule RIN1 (Ras pathway component Ras and Rab interactor 1). Hu et al. (2005) showed that RIN1 activates ABL2 by interaction of a poly-proline motif with the ABL SH3 domain. Domain binding results in phosphorylation of C-terminal RIN1 Y36 by ABL and thus creates a binding site for the ABL SH2 domain. Hence, ABL kinase activity is stimulated by conformational changes suspending auto-inhibitory mechanisms. Tyrosine 36 of RIN1 was motif predicted to be targeted by ABL2 and YES1. While ABL2 is known to modify Y36 of RIN1 and thereby undergoes activation in *trans*, YES1 targeting of this residue may additionally activate ABL kinases via the RIN1 signaling adapter.

Neuronal migration during brain development is controlled by signaling molecules such as CDK5 as mentioned above and in particular by the Reelin-Disabled-1 (Dab1) signaling pathway involving CRK. The Reelin gene encodes a secreted glycoprotein which can bind to very low density lipoprotein receptor (VLDLR) and low density lipoprotein receptor-related protein 8, apolipoprotein e receptor (APOER2) receptors. Reelin binding induces tyrosine phosphorylation of adaptor protein Dab1 resulting in accurate neuronal positioning as reviewed by Gao and Godbout (2013). Dab1 contains a N-terminal phospho-tyrosine binding domain that binds to Reelin receptors and an internal tyrosine rich region consisting of five conserved tyrosine residues Y185, Y198, Y200, Y220, and Y232 of which

Y198, Y220, and Y232 were identified to be phosphorylated upon stimulation by Reelin (Gao and Godbout, 2013). Phosphorylation of Dab1 recruits SH2 domain containing proteins and hence activates downstream signaling pathways dependent on concomitantly phosphorylated pairs of residues. Phosphorylated Y185 and Y198 of Dab1 enables binding of the SH2 domains of SFKs, phosphatidylinositide-3-kinase (PI3K), and suppressor of cytokine signaling (SOCS) whereas Y220 and Y232 modification facilitate binding of ABL, CRK and CRKL SH2 domains (Gao and Godbout, 2013). SFKs, and in particular FYN (Arnaud et al., 2003), were determined as mainly responsible kinases for Dab1 modifications. Most SH2 binding proteins require only a single Dab1 tyrosine residue to be phosphorylated for binding with the exception of CRK and CRKL which bind Dab1 only via simultaneously phosphorylated Y220 and Y232 (Ballif et al., 2004). Both Y220 and Y232 were predicted via sequence motif to be targeted by ABL2. This prediction recapitulated a report by Pramatarova et al. (2003) who performed an *in vitro* binding assay using lysates of transfected HEK293T cells. GST-fused Dab1 was phosphorylated by mouse ABL on both Y220 and Y232 and subsequently tested for binding to a downstream adapter molecule NCK beta. NCK beta and SH2 containing platelet-activating factor acetylhydrolase 1b, regulatory subunit 1 (45kDa) (LIS1) can bind Dab1 on either phosphorylated Y220 and Y232 or either phosphorylated Y220 and Y198, respectively (Ballif et al., 2004). Interestingly, Dab1 Y220 was furthermore motif predicted to be targeted by FER and FRK suggesting a potential role of these NRTKs in neuronal migration as phosphorylation of Y220 may trigger or may support activation of downstream signaling by NCK beta, LIS1, and CRK proteins.

In summary, 1388 human pY-sites accounting for approximately ten percent of all reported human pY-sites were assigned to NRTKs by sequence motif scoring and some predictions were validated experimentally and backed by literature. While most motif the predictions are specific kinase-substrate pairs, also overlapping NRTK targeting was prognosticated for single and multiple sites on putative substrates. Literature reported substrate targeting involved mainly single or very few kinases for a given pY-site. Therefore, motif-based scoring as presented here could identify additional NRTKs potentially involved in the reported NRTK-driven cellular processes. Regulation of CDKs and their regulators by the SFK family was suggested providing an example for motif predicted targeting of related groups of substrates by NRTK families or related PTKs. Furthermore, nine of 16 NRTKs were themselves predicted to possess NRTK motif matching pY-sites from almost all NRTK families indicating the complexity of inter-kinase regulation. Finally, modification of CRK, RIN1, and related proteins was discussed as examples for predictions that are strongly supported by current literature. Thereby, functional consequences of tyrosine phosphorylation and the power of the motifs generated here to identify putative NRTKs which are involved were exemplified.

5. Conclusion

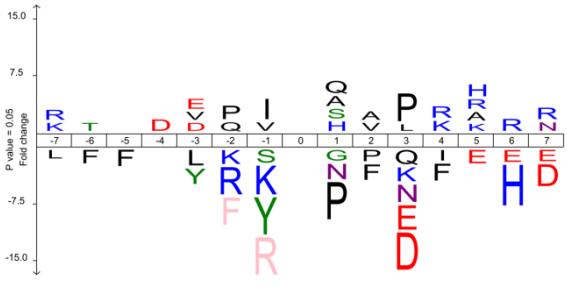
Oncogenic tyrosine kinase signaling is regarded as a hallmark of cancer and disease and constitutes a major field of molecular biomedicine. The determination of NRTK specificity is essential for understanding tyrosine signaling and serves as a basis for drug development including NRTK inhibitors. A major challenge in NRTK characterization using mammalian cells constitutes overlapping substrate targeting together with tissue-specific NRTK expression. Therefore, even though thousands of mammalian phosphorylation sites are continuously determined by proteomic approaches only a minor fraction are assigned to particular NRTKs. Previous efforts focused on *in vitro* determination of individual NRTK primary sequence specificities which may explain only up to 20 percent of all KSRs. Contextual information such as 3D substrate structure, protein complex formation, and subcellular localization appears to be another major specificity determinant. Employing yeast as the best-characterized model system for mimicking the cellular environment of mammalian cells overcomes some previous limitations by analyzing individual NRTK targeting *in vivo* in a lower eukaryote providing fully-folded substrates in a highly-crowded environment in a PTK signaling background-free manner. NRTK specificity was surprisingly well maintained in yeast as determined by systematic analysis of biochemical and phylogenetic properties of NRTK targeted yeast tyrosine residues. Half of all cytoplasmic PTKs showed activity in yeast and were tested individually under the same cellular conditions. Thus, large yeast substrate sets were generated providing a unique dataset to analyze NRTK targeting overlap and to generate novel, high-performing linear sequence motifs naturally containing contextual information. The dataset enabled observation of direct PPI network targeting suggesting that specificity does not necessarily have to be encoded in substrates themselves. The specificities determined in yeast were transferred to human via site conservation and by motif-based scoring of human pY-sites. Therefore, a total of 1451 tyrosine phosphorylation sites were assigned to single and multiple NRTKs across all NRTK families – comprising around 10 percent of all measured instances in human cells. Selected predictions were validated experimentally by an *in vitro* kinase assay and several predictions were supported by current literature. Among 63 measured pY-sites conserved in human, almost all metabolic enzymes involved in glycolysis (and not in oxidative phosphorylation) were identified as NRTK targets. These findings suggest a previously unknown extent of cytoplasmic PTK signaling-dependent regulation of cellular respiration. The hypothesis derived is that tyrosine phosphorylation may in part drive the Warburg effect observed in proliferating cells. Motif-based predictions enabled confident matching of hundreds of substrates to specific NRTKs whereas overlapping predictions suggested targeting of single proteins or tyrosine residues, or targeting of groups of substrates by multiple NRTKs. Thus, motif-based targeting of cell cycle regulators by SFKs, FER, and FRK was predicted which partly recapitulates and extends previous

knowledge. Moreover, the complexity of inter-kinase regulation was illustrated by identification of many NRTK motif matching sites on NRTK protein sequences. Furthermore, previously reported regulatory mechanisms in CRK adapter molecule signaling were discussed as an example for functional consequences of motif predicted tyrosine phosphorylation. All predictions of CRK modification by ABL2 kinase were validated by research reports and the involvement of additional NRTKs was proposed. Overall, expression of NRTKs in intact yeast to study signaling specificity is a novel and successful concept addressing both the primary sequence of yeast substrates and the cellular context provided. Prediction of KSRs based on yeast data may aid further experimental investigations to determine how NRTK signaling is globally regulating cellular processes in normal and proliferating cells.

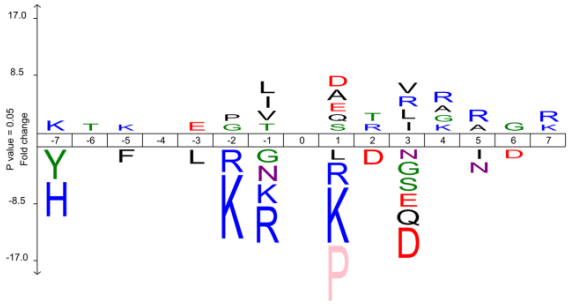
6. Appendix

6.1. Linear sequence motifs

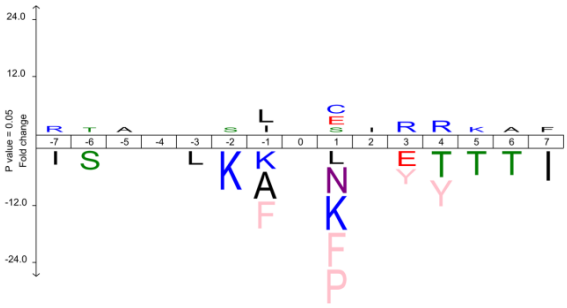
ABL2 motif



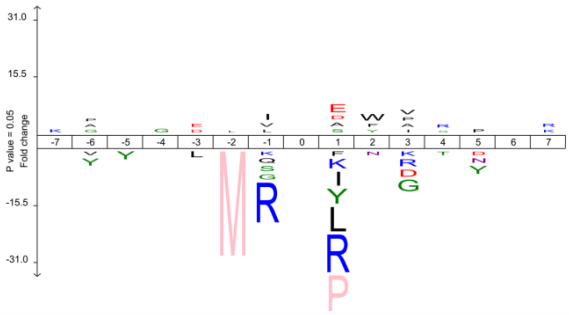
BLK motif



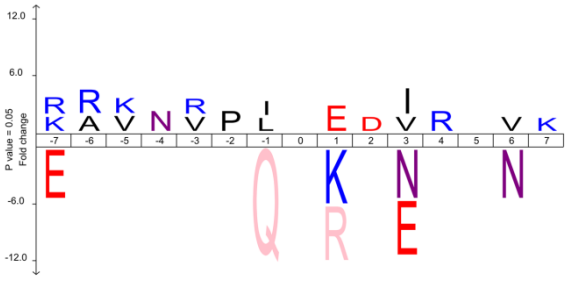
BMX motif



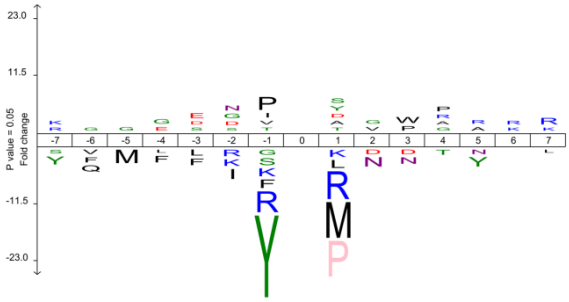
FER motif



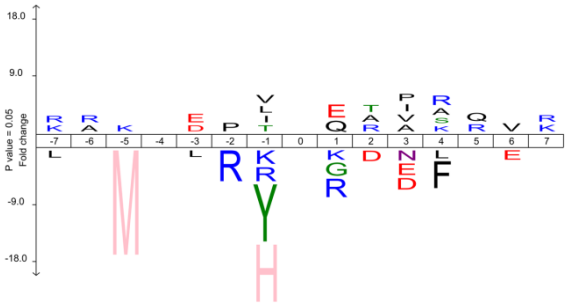
FES motif



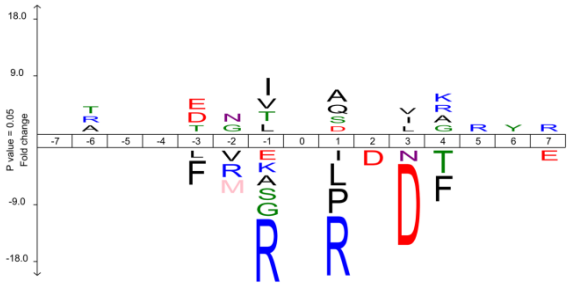
FGR motif

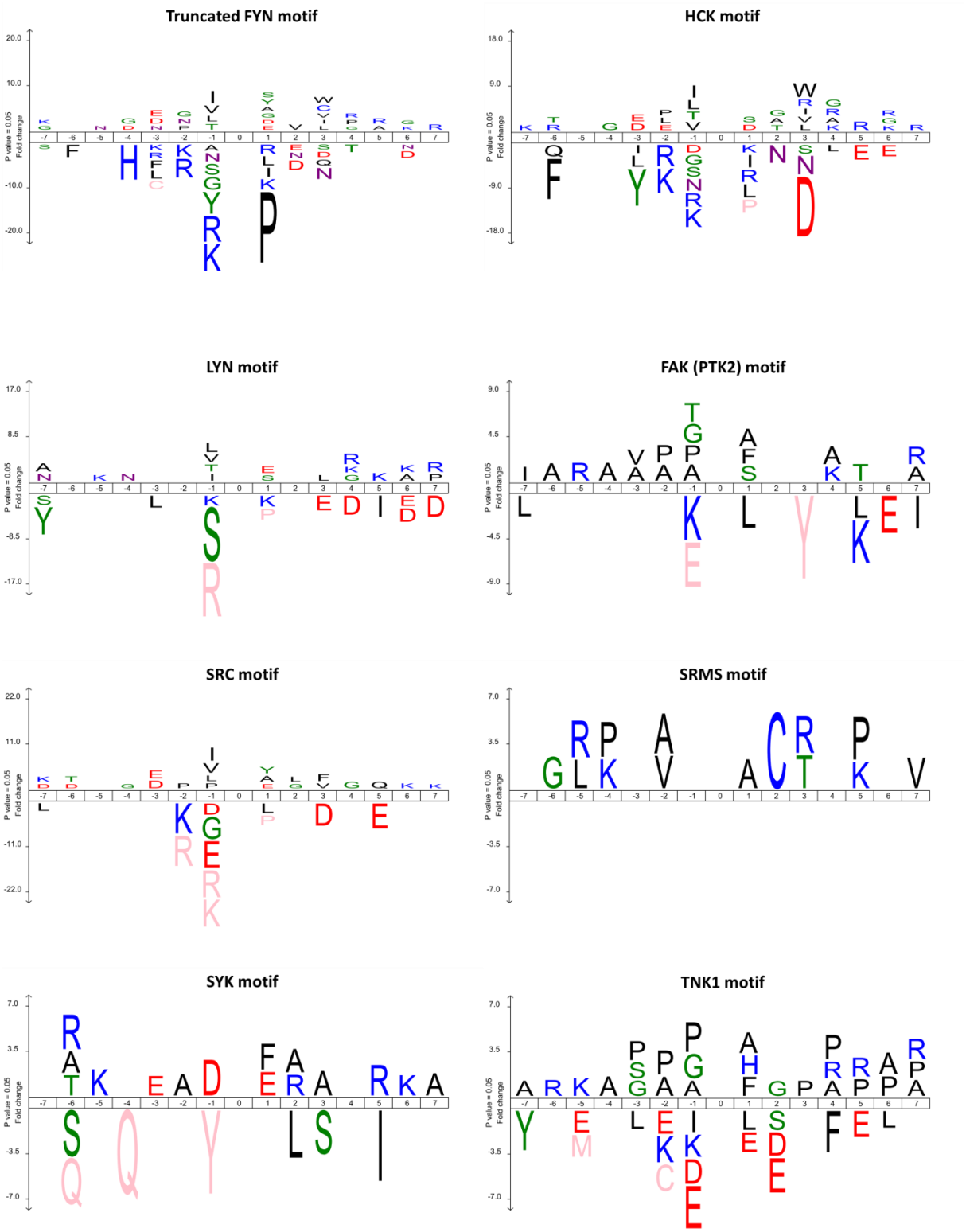


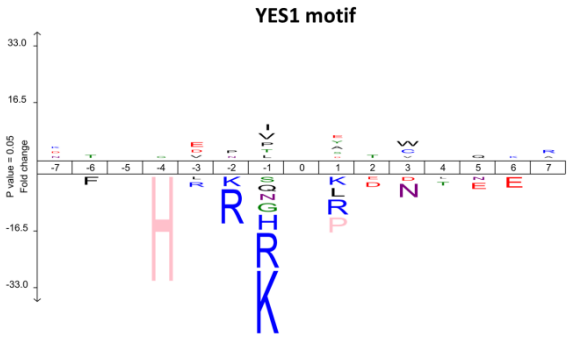
FRK motif



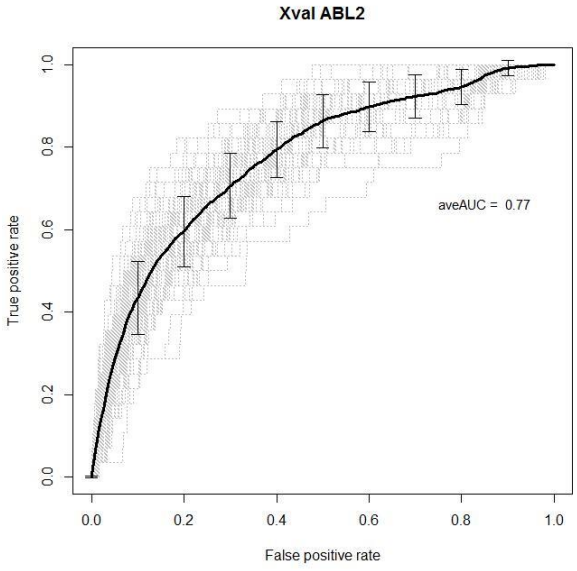
FYN motif

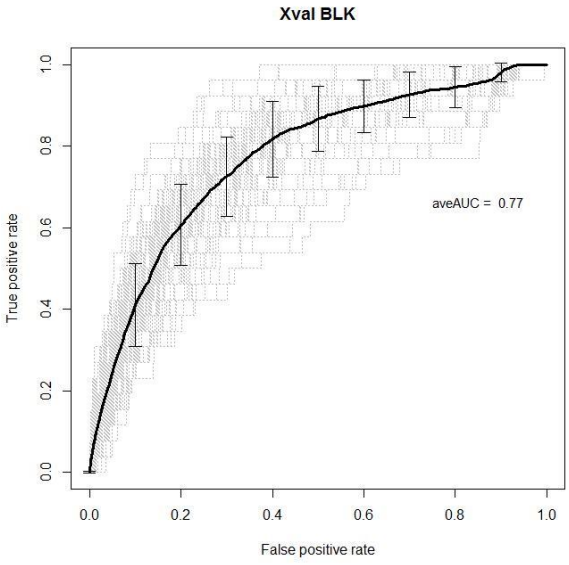


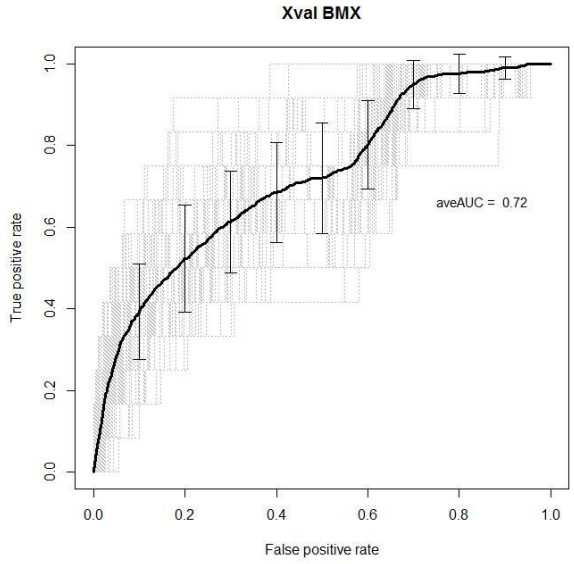

Truncated FYN motif
HCK motif
LYN motif
FAK (PTK2) motif
SRC motif
SRMS motif
SYK motif
TNK1 motif

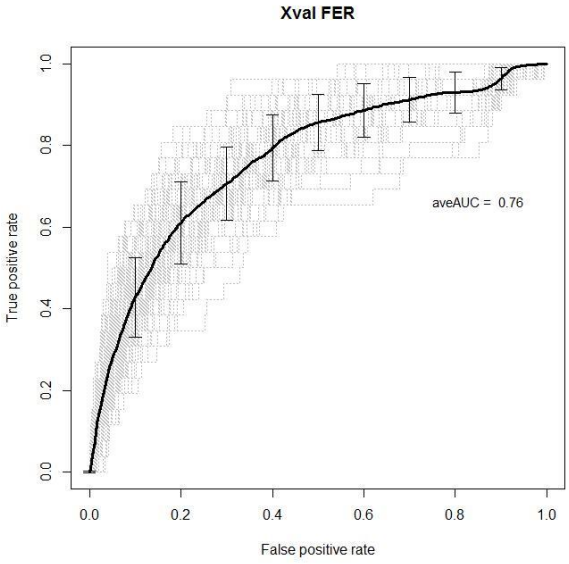


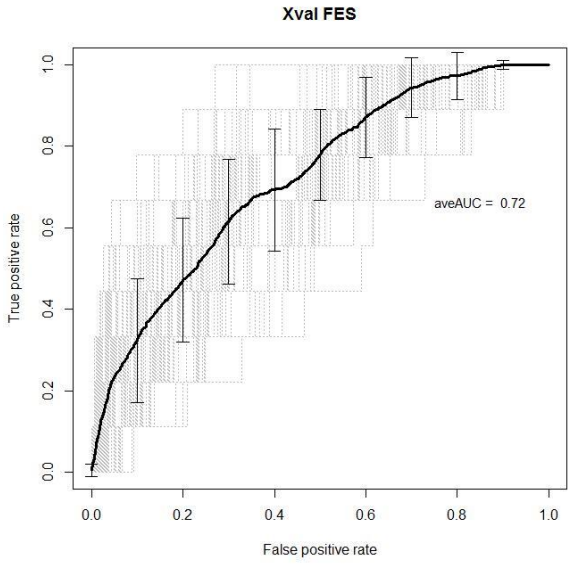
6.2. ROC cross-validations

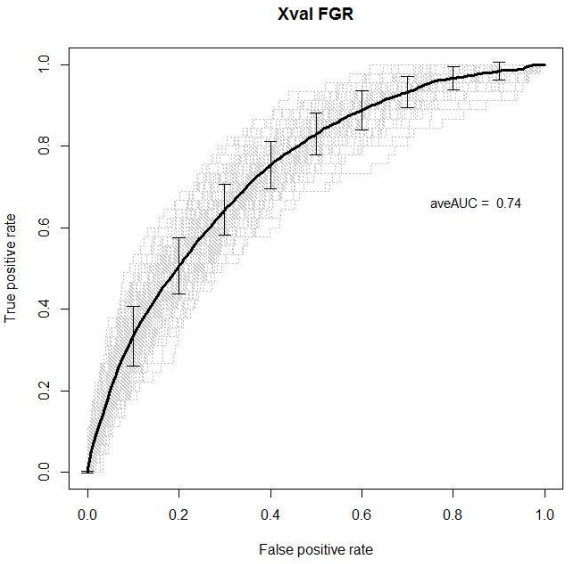


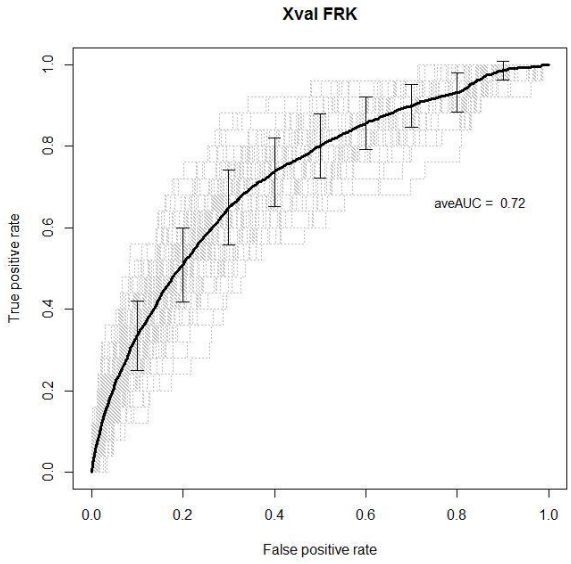


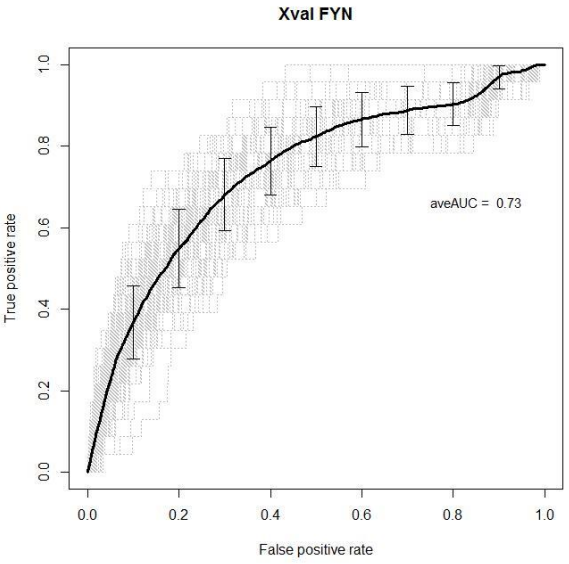


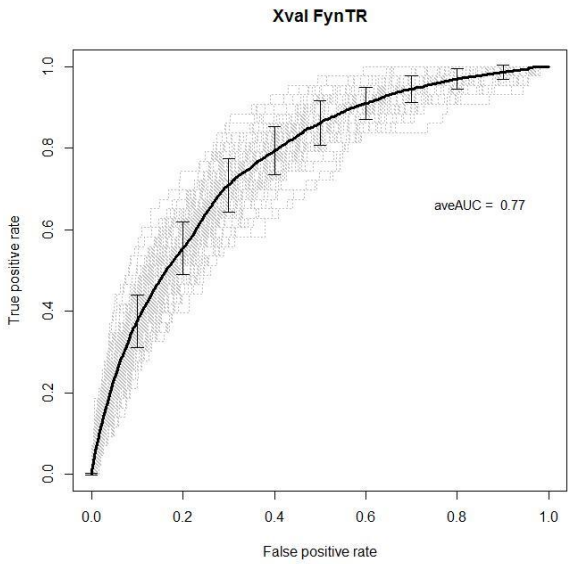


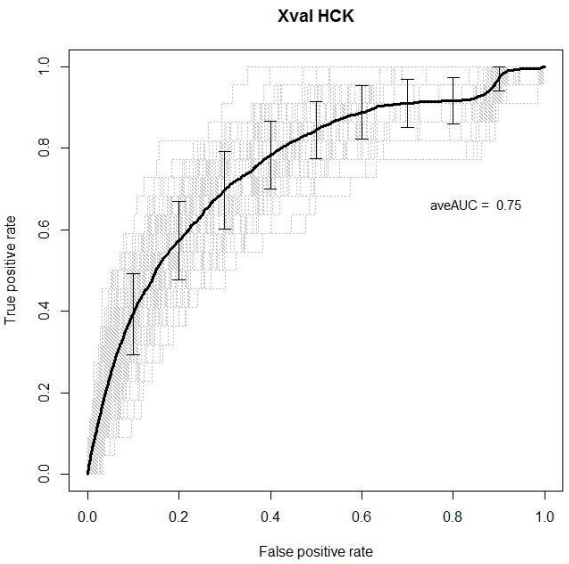

Xval FES
True positive rate
False positive rate
aveAUC = 0.72
0.0
0.2
0.4
0.6
0.8
1.0

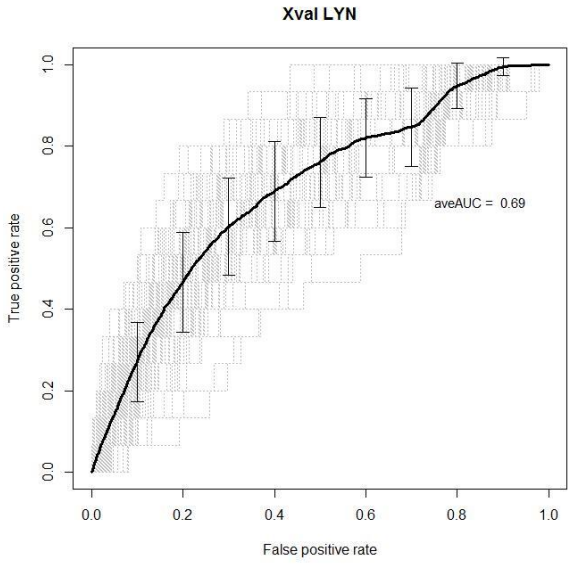

Xval FGR
True positive rate
False positive rate
aveAUC = 0.74
0.0
0.2
0.4
0.6
0.8
1.0

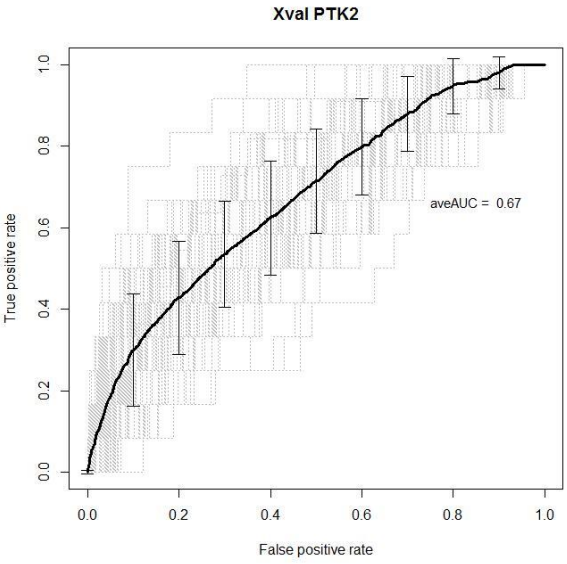

Xval FRK
True positive rate
False positive rate
aveAUC = 0.72
0.0
0.2
0.4
0.6
0.8
1.0

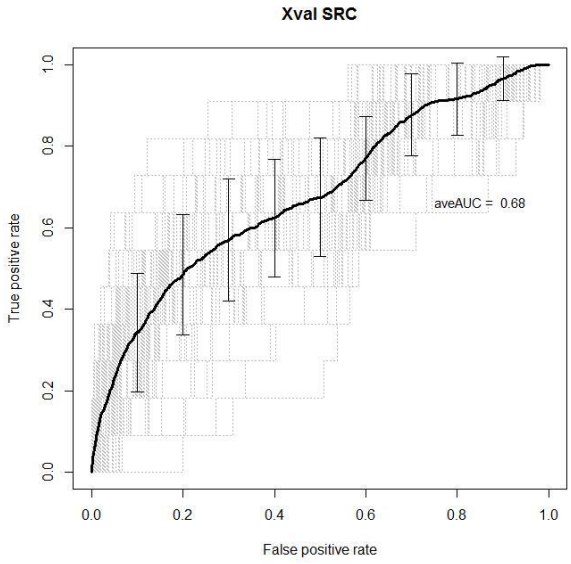

Xval FYN
True positive rate
False positive rate
aveAUC = 0.73
0.0
0.2
0.4
0.6
0.8
1.0

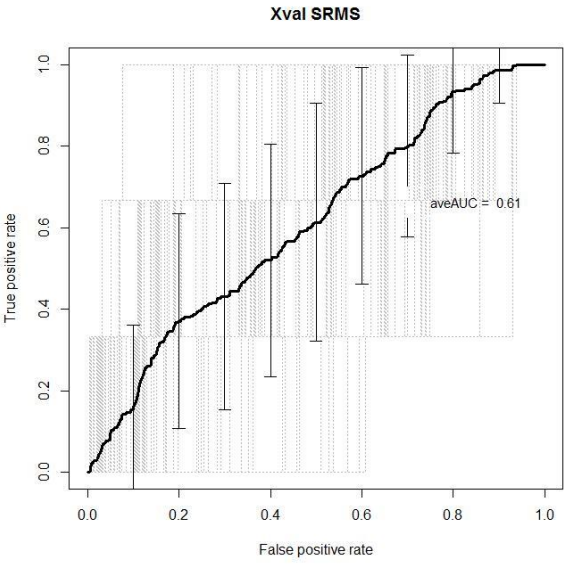

Xval FynTR
True positive rate
False positive rate
aveAUC = 0.77
0.0
0.2
0.4
0.6
0.8
1.0

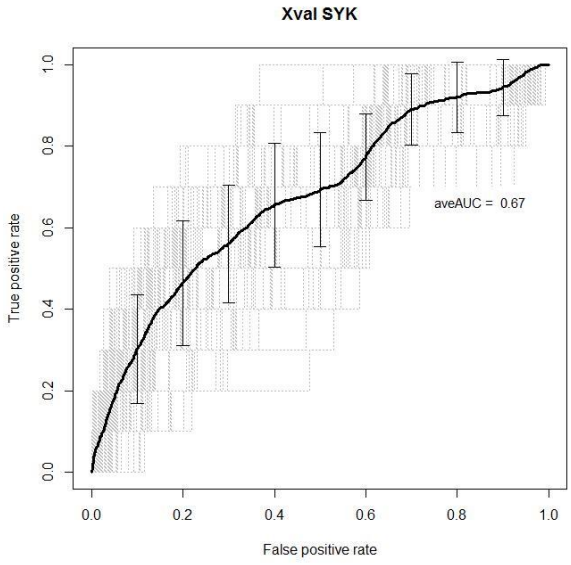

Xval HCK
True positive rate
False positive rate
aveAUC = 0.75
0.0
0.2
0.4
0.6
0.8
1.0

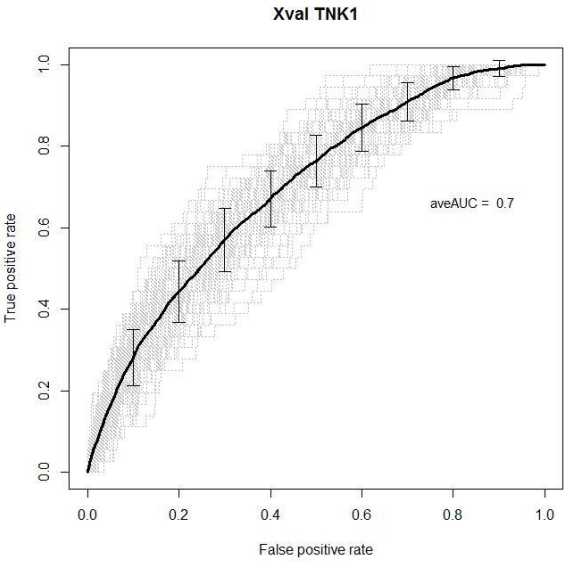

Xval LYN
True positive rate
False positive rate
aveAUC = 0.69
0.0
0.2
0.4
0.6
0.8
1.0

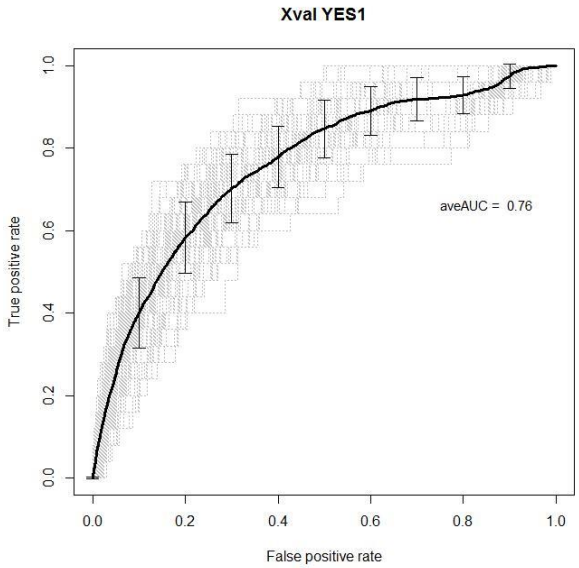

Xval PTK2
True positive rate
False positive rate
aveAUC = 0.67
0.0
0.2
0.4
0.6
0.8
1.0


Xval SRC
True positive rate
False positive rate
aveAUC = 0.68
0.0
0.2
0.4
0.6
0.8
1.0

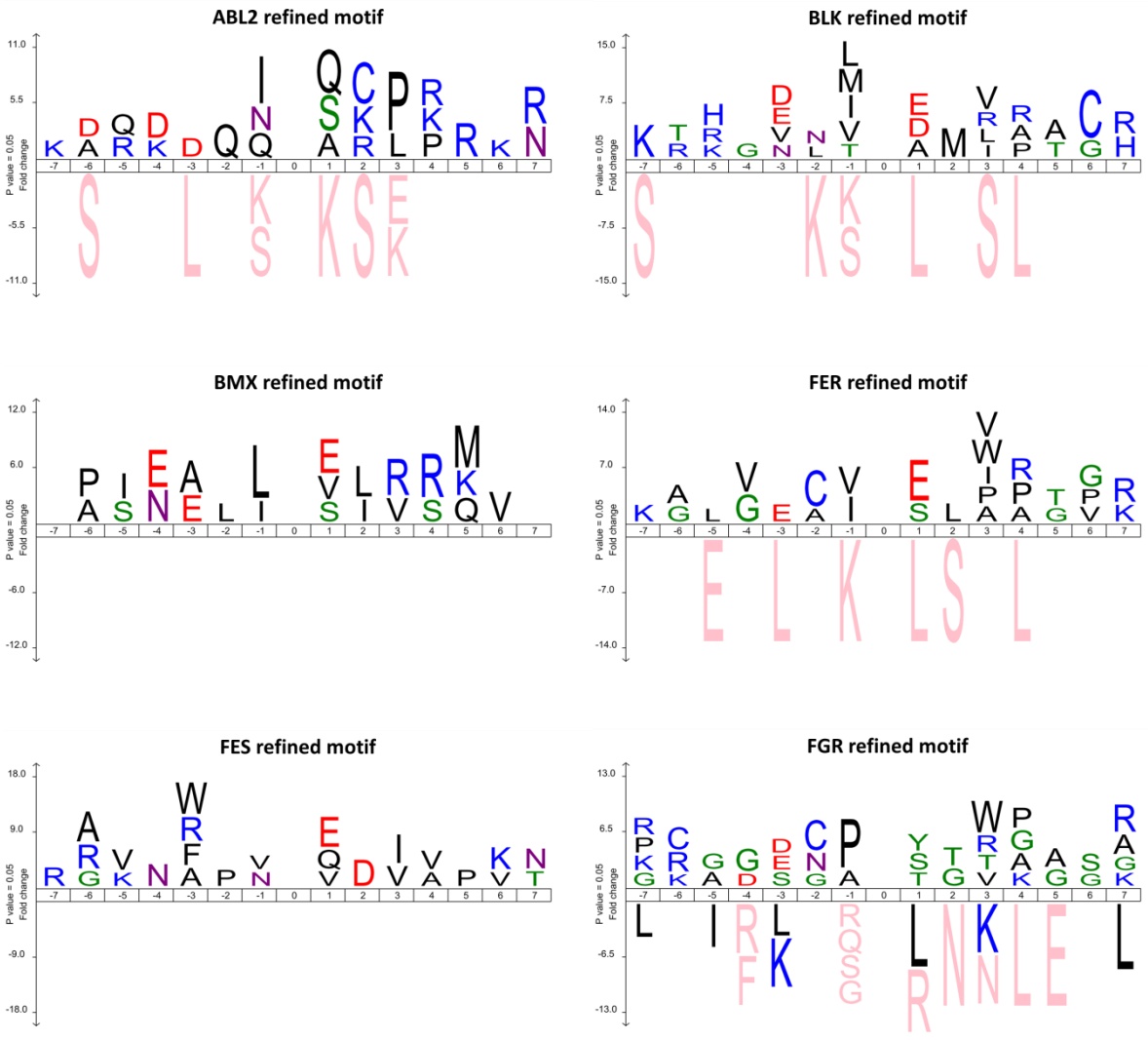

Xval SRMS
True positive rate
False positive rate
aveAUC = 0.61
0.0
0.2
0.4
0.6
0.8
1.0


Xval SYK
True positive rate
False positive rate
aveAUC = 0.67
0.0
0.2
0.4
0.6
0.8
1.0

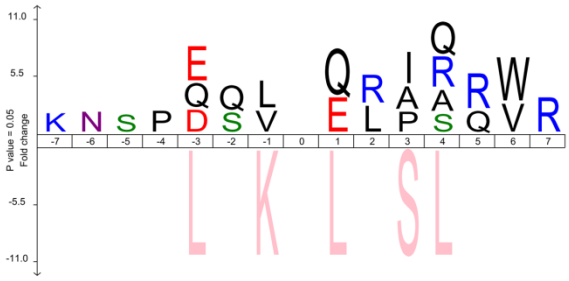

Xval TNK1
True positive rate
False positive rate
aveAUC = 0.7
0.0
0.2
0.4
0.6
0.8
1.0



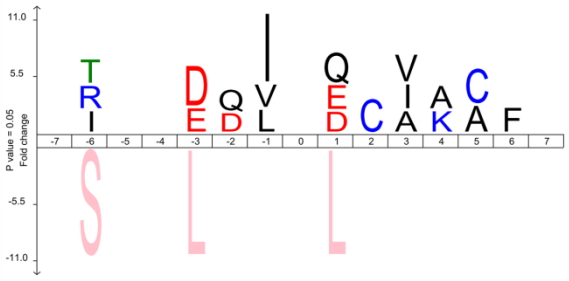
6.3. Refined linear sequence motifs



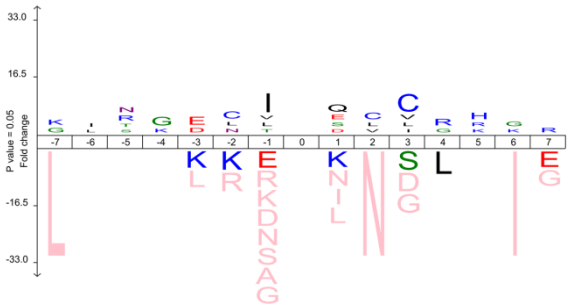
FRK refined motif



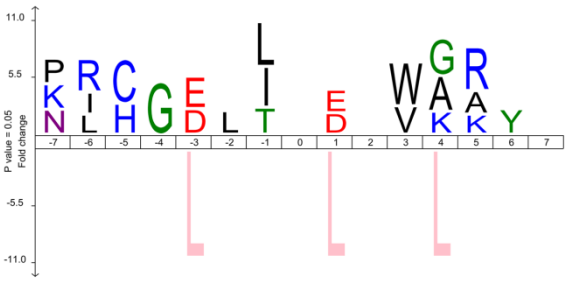
FYN refined motif



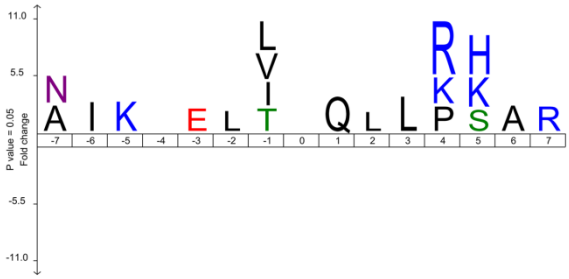
Truncated FYN refined motif



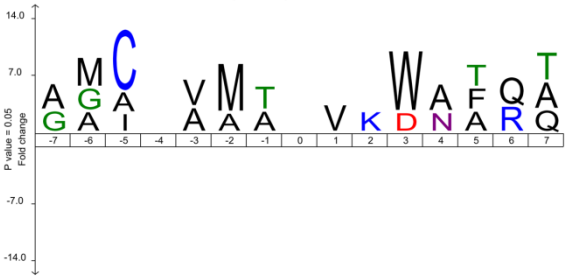
HCK refined motif



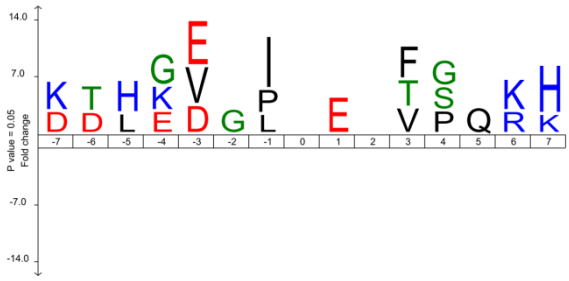
LYN refined motif



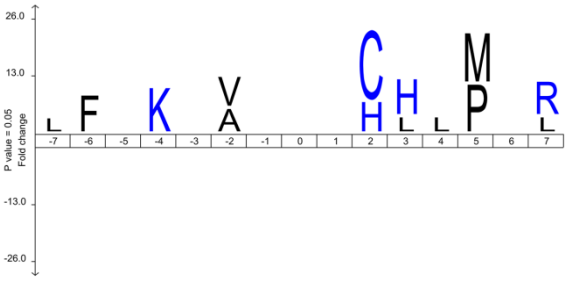
FAK (PTK2) refined motif



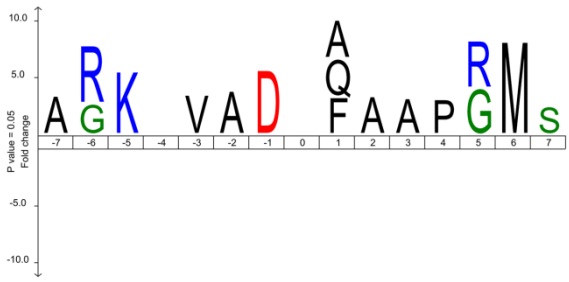
SRC refined motif



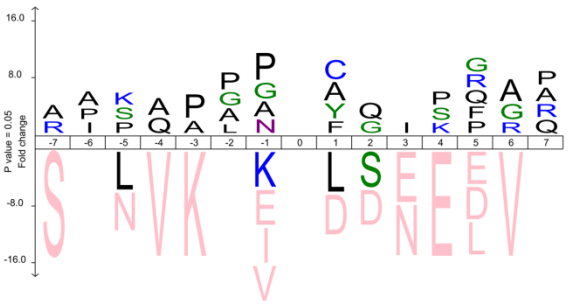
SRMS refined motif

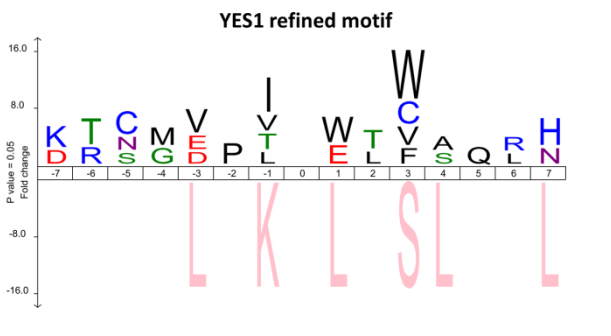


SYK refined motif

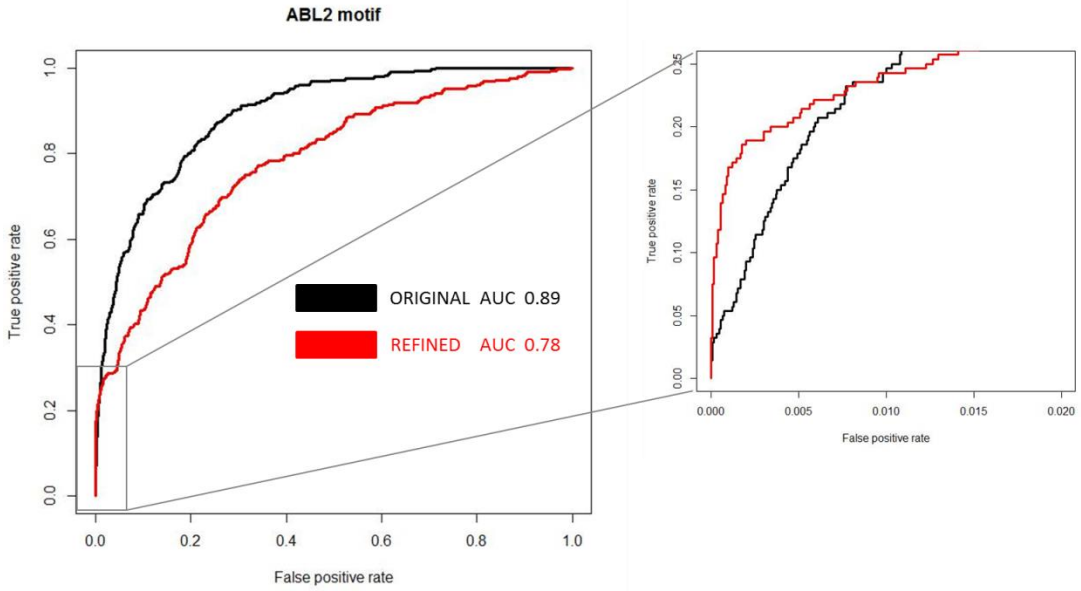


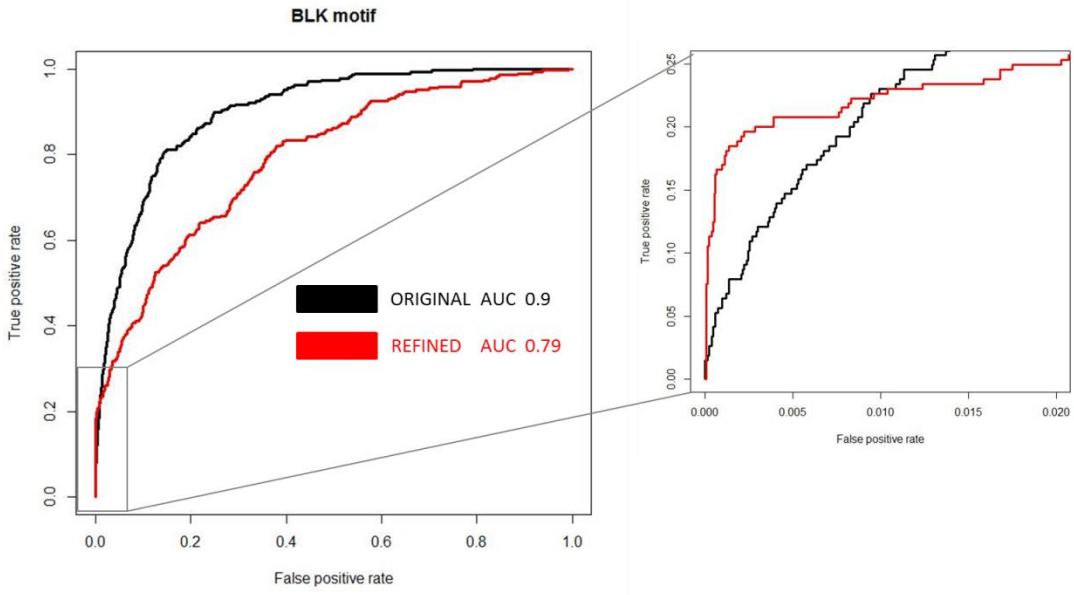
TNK1 refined motif

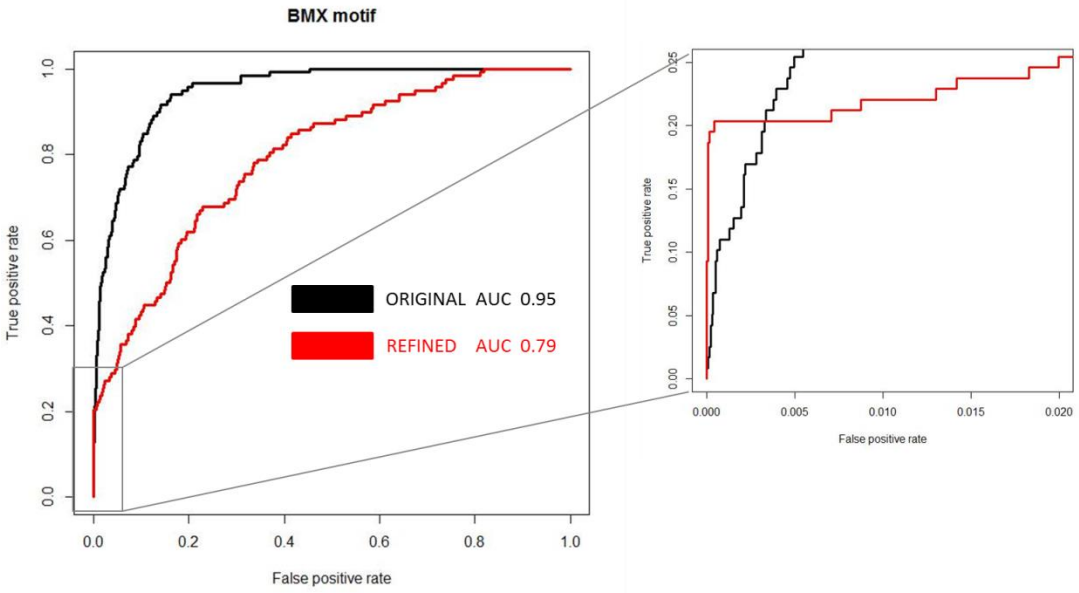


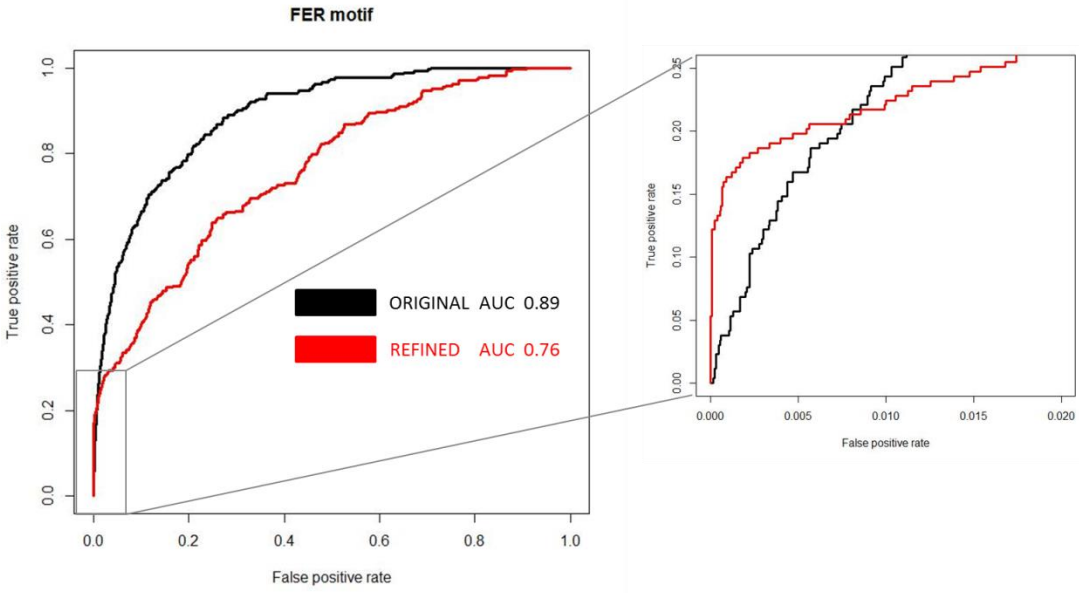


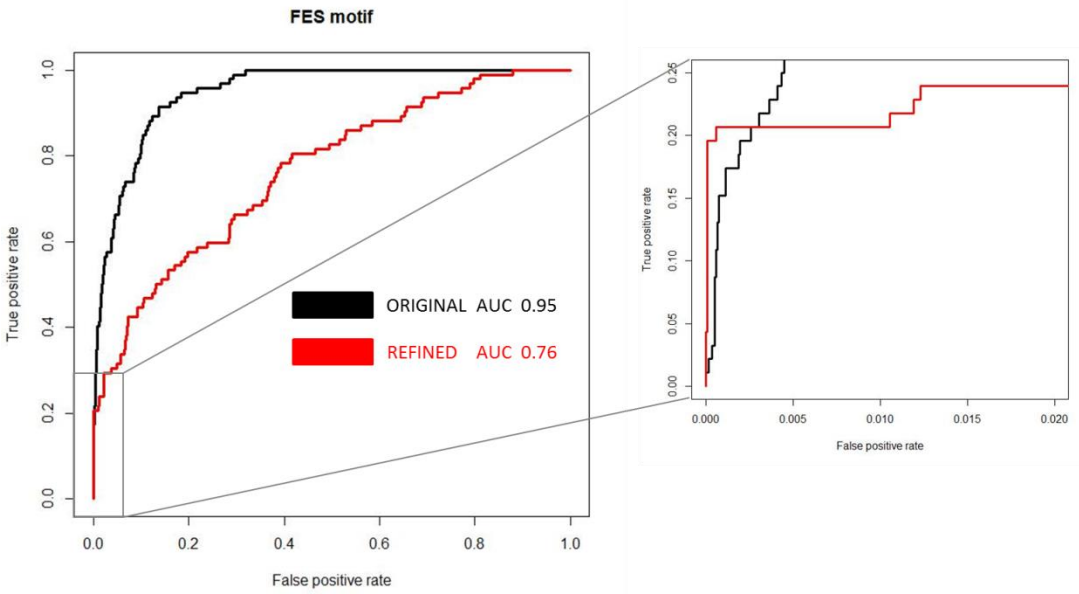
6.4. ROC performance comparisons original motifs versus refined motifs

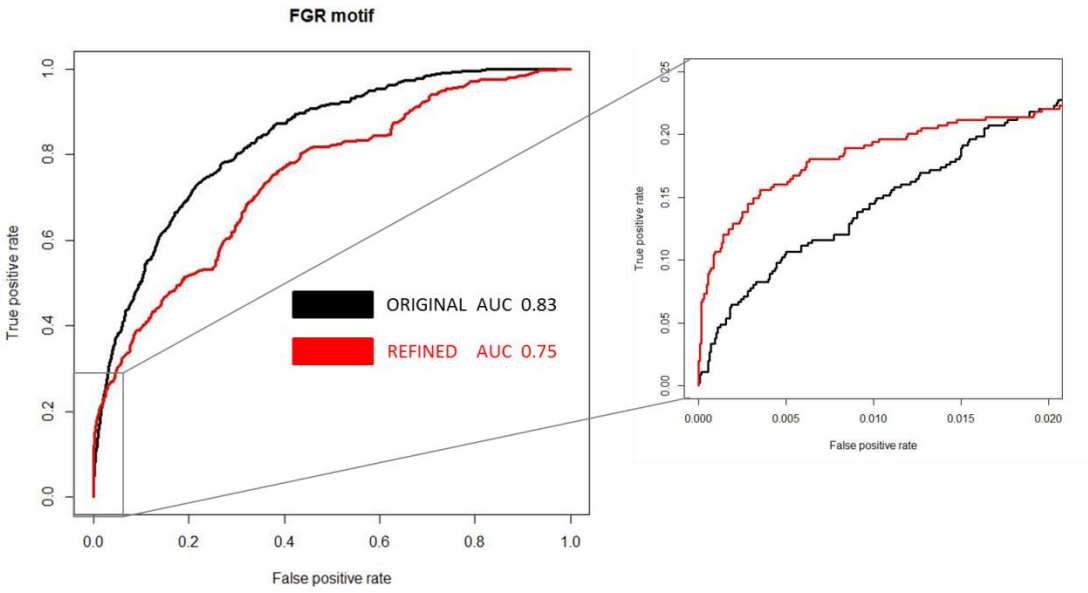


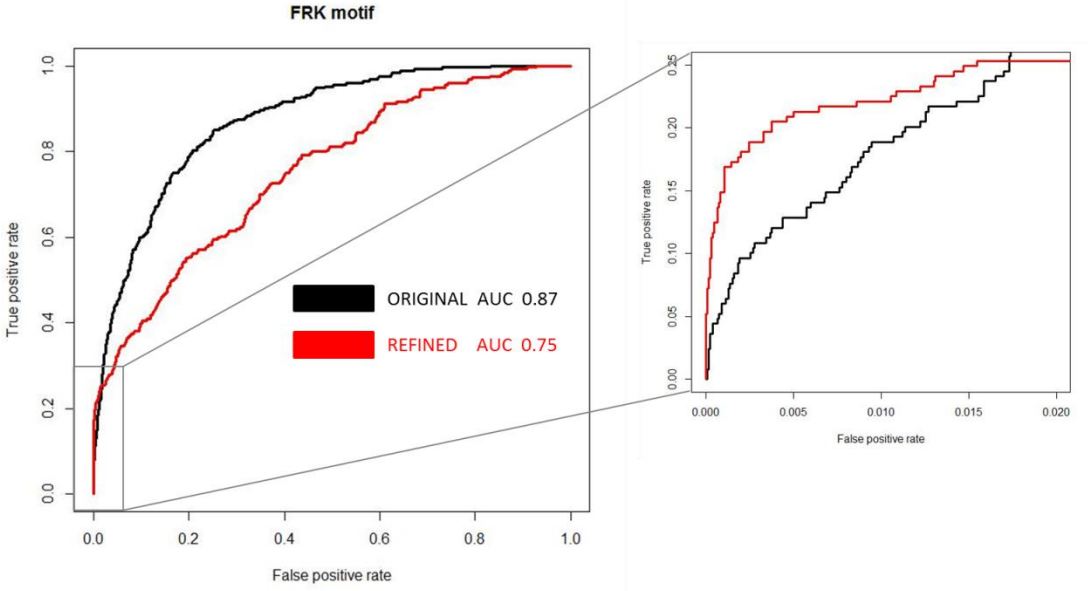


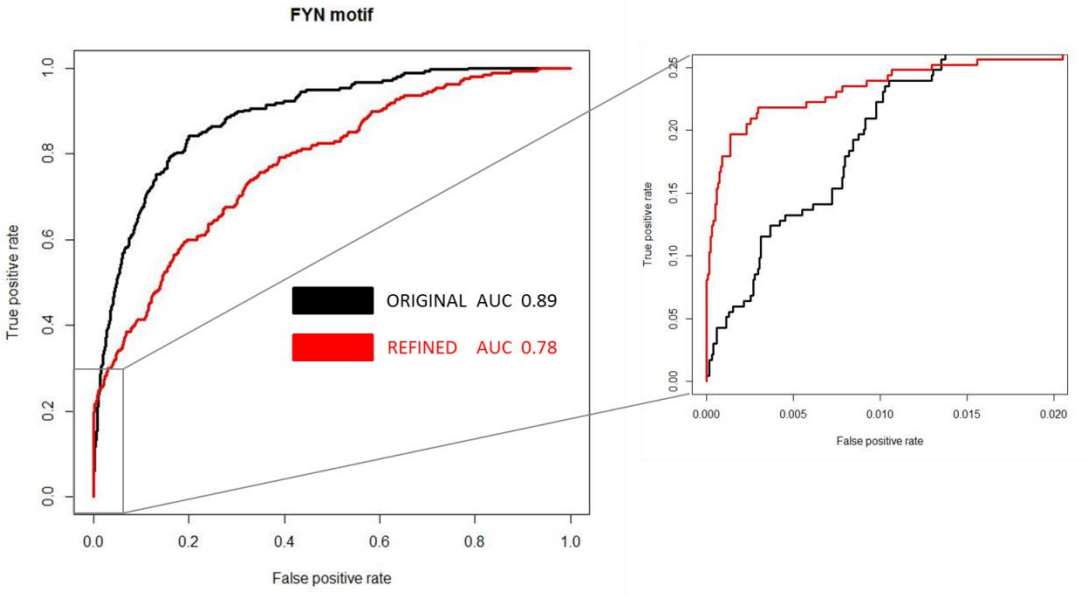

BMX motif
True positive rate
False positive rate
ORIGINAL AUC 0.95
REFINED AUC 0.79

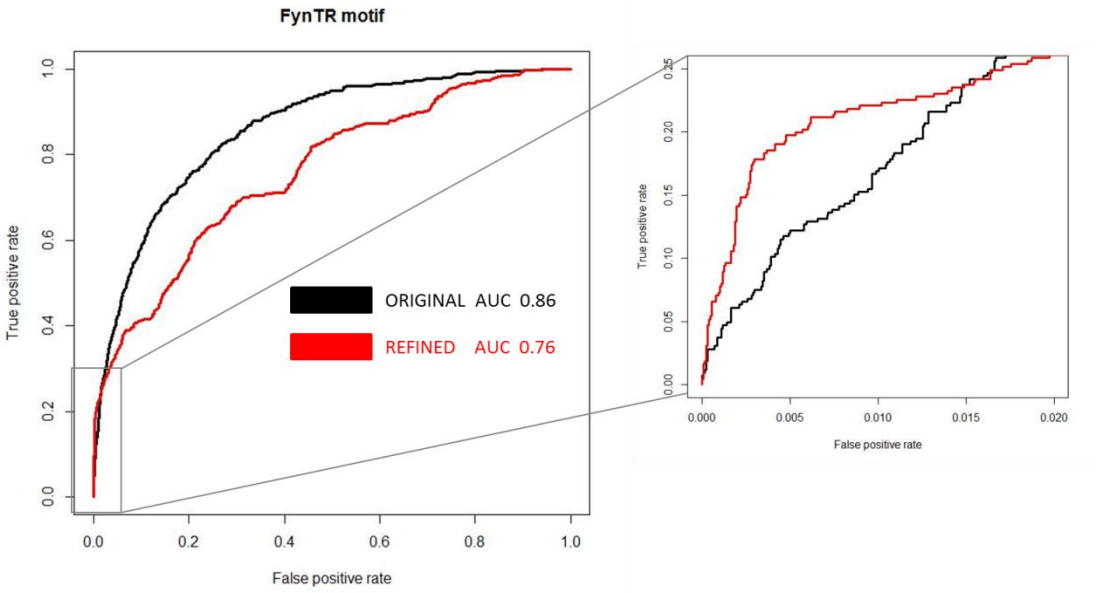

FER motif
True positive rate
False positive rate
ORIGINAL AUC 0.89
REFINED AUC 0.76

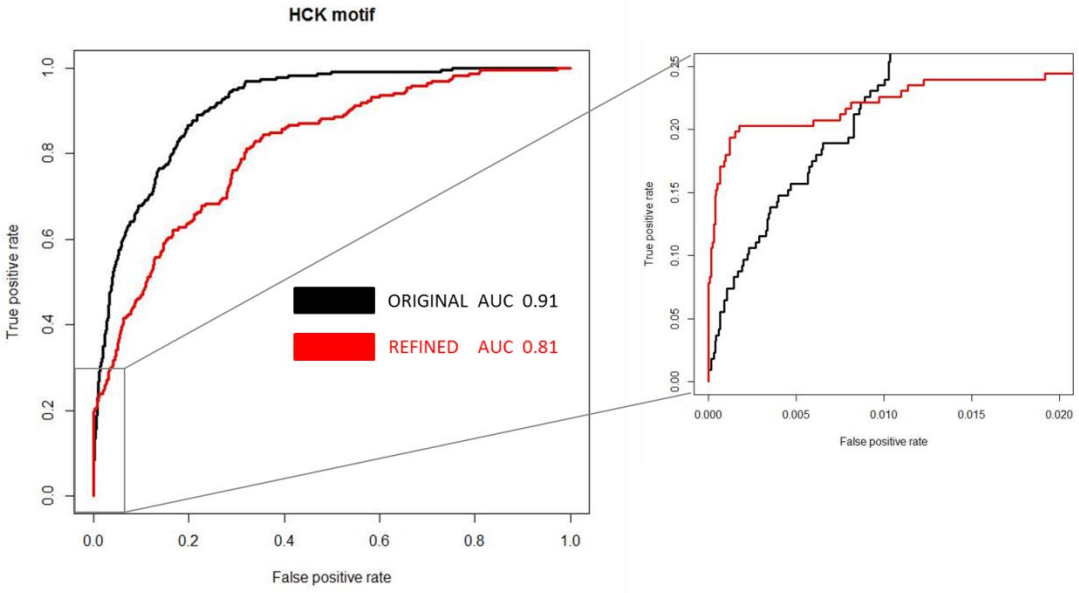

FES motif
True positive rate
False positive rate
ORIGINAL AUC 0.95
REFINED AUC 0.76

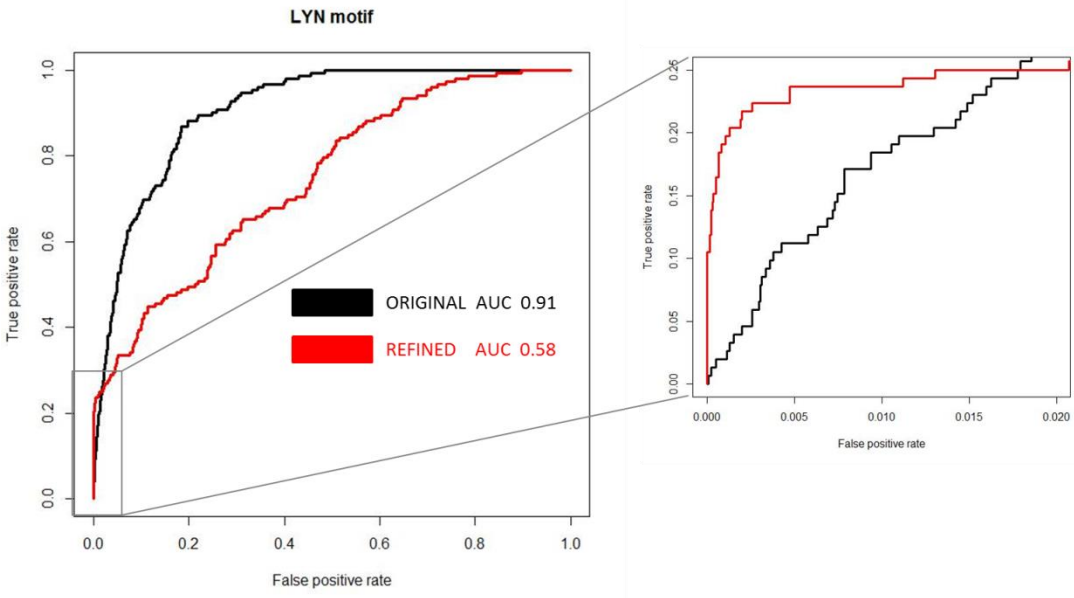

FGR motif
True positive rate
False positive rate
ORIGINAL AUC 0.83
REFINED AUC 0.75

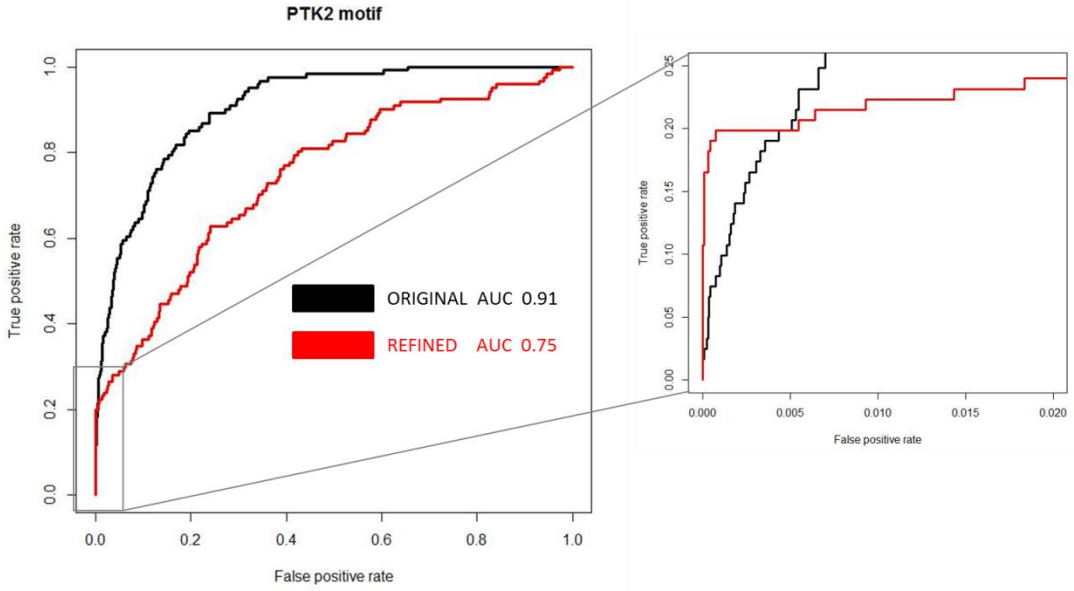

FRK motif
True positive rate
False positive rate
ORIGINAL AUC 0.87
REFINED AUC 0.75

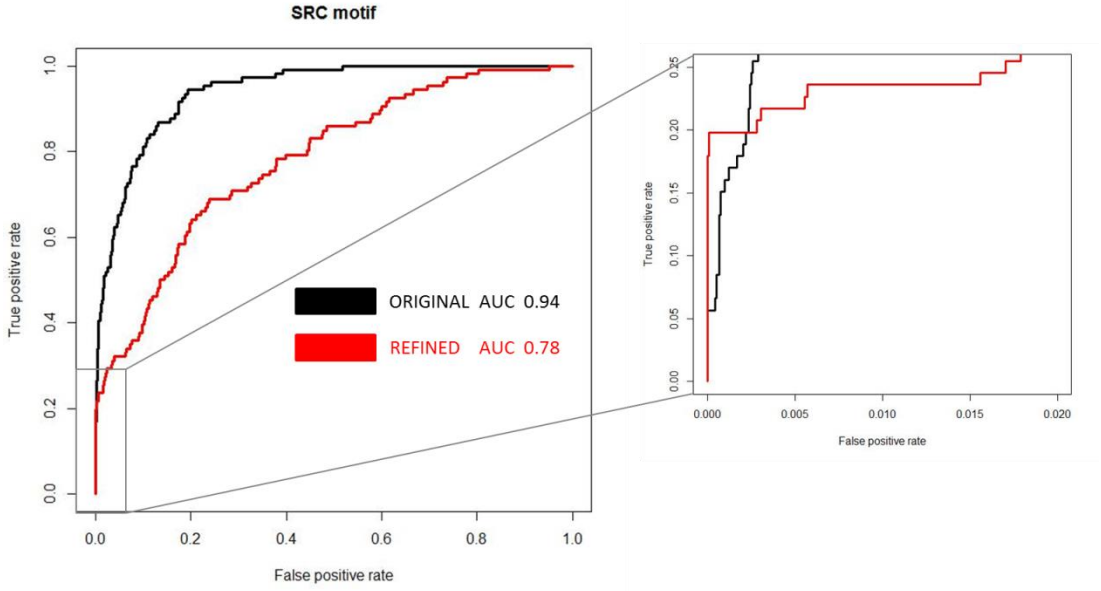

FYN motif
True positive rate
False positive rate
ORIGINAL AUC 0.89
REFINED AUC 0.78

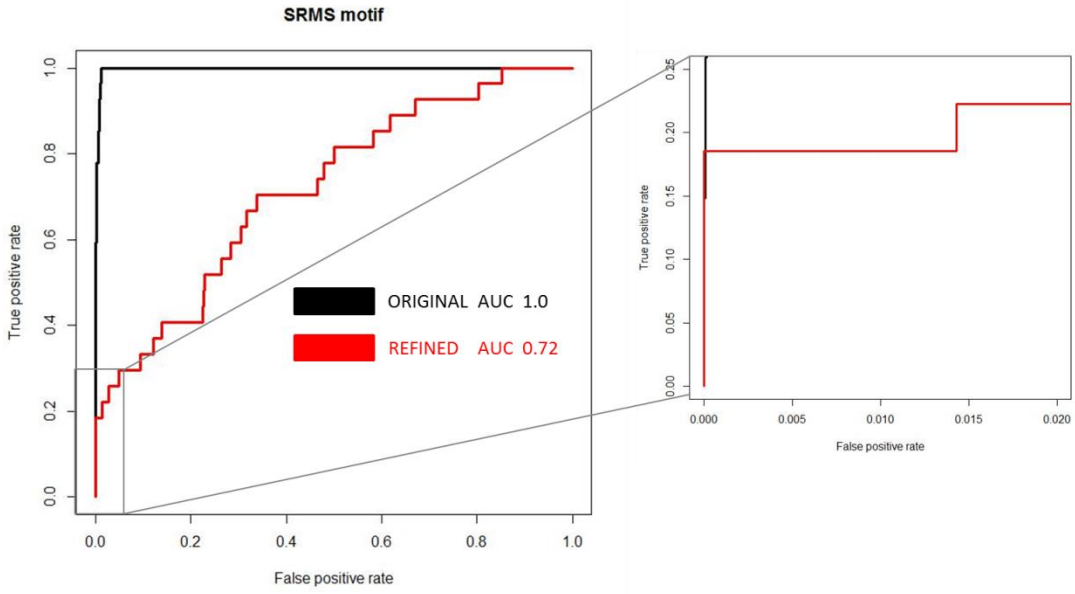

FynTR motif
True positive rate
False positive rate
ORIGINAL AUC 0.86
REFINED AUC 0.76

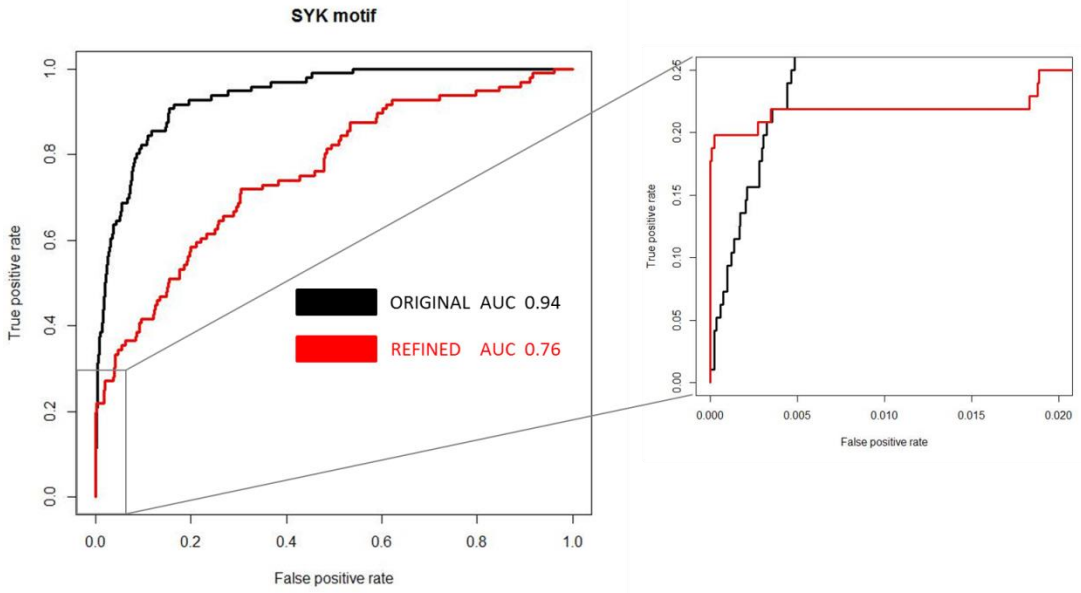

HCK motif
True positive rate
False positive rate
ORIGINAL AUC 0.91
REFINED AUC 0.81

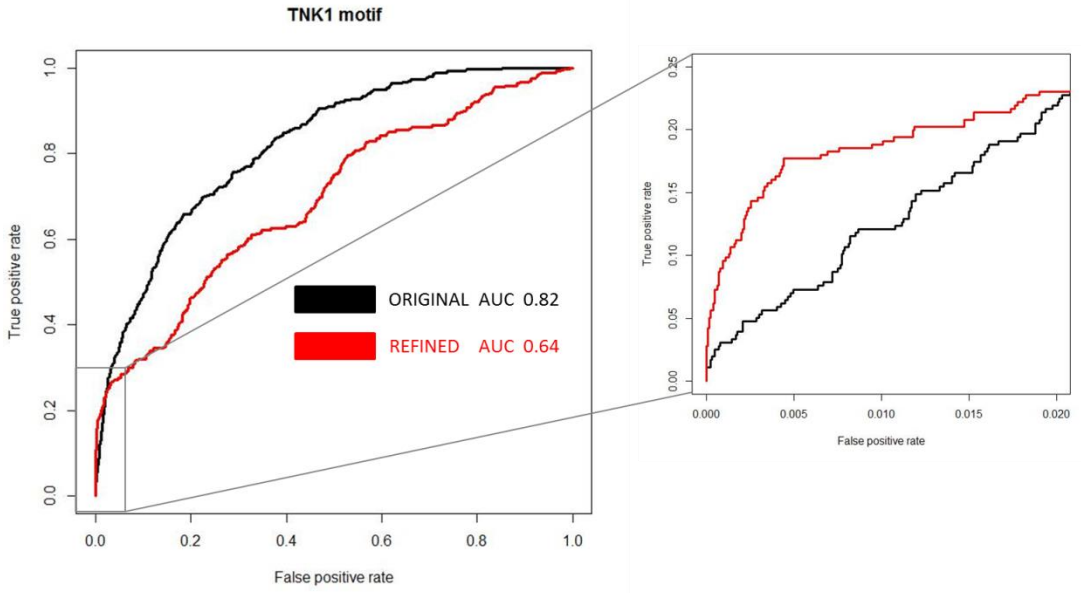

LYN motif
True positive rate
False positive rate
ORIGINAL AUC 0.91
REFINED AUC 0.58

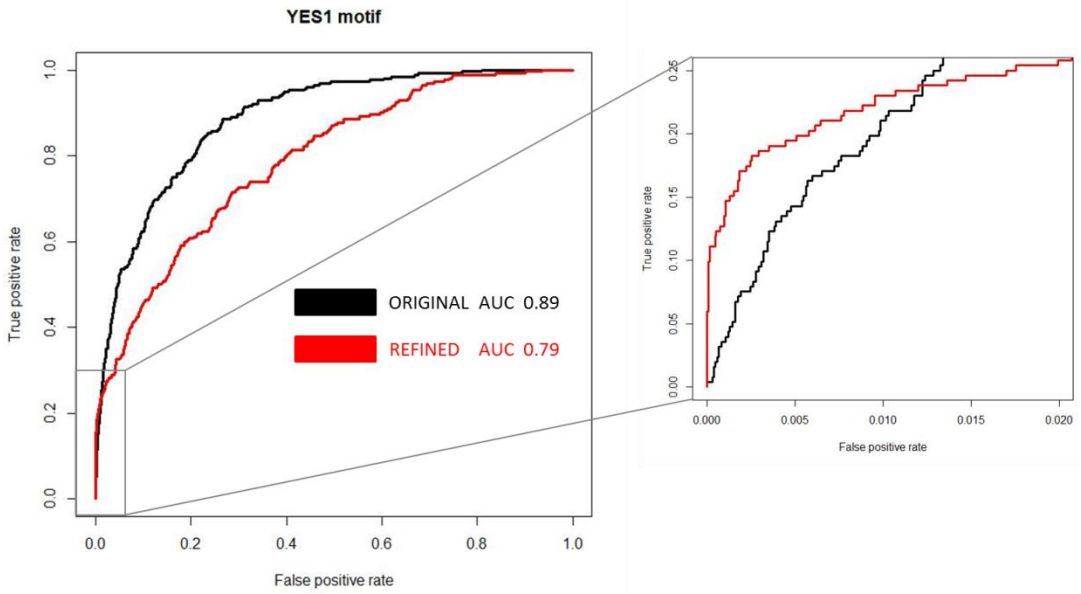

PTK2 motif
True positive rate
False positive rate
ORIGINAL AUC 0.91
REFINED AUC 0.75


SRC motif
True positive rate
False positive rate
ORIGINAL AUC 0.94
REFINED AUC 0.78

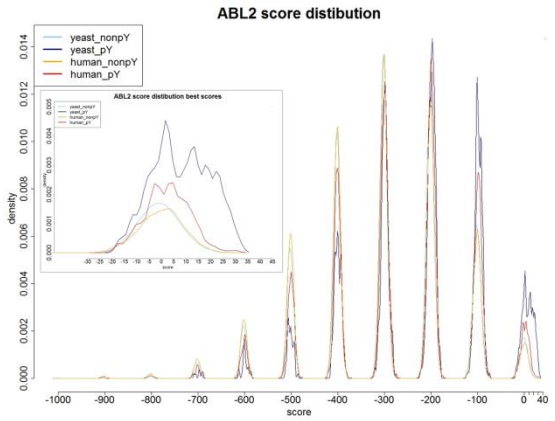

SRMS motif
True positive rate
False positive rate
ORIGINAL AUC 1.0
REFINED AUC 0.72

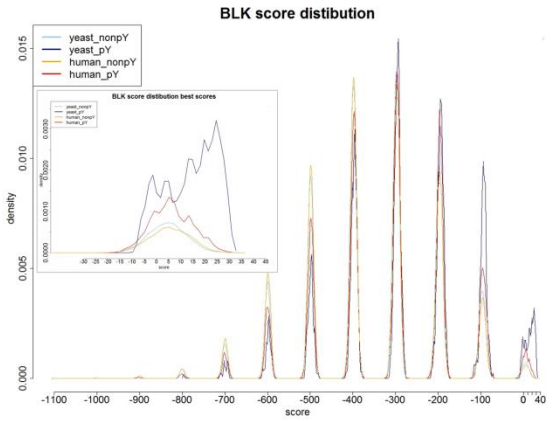

SYK motif
True positive rate
False positive rate
ORIGINAL AUC 0.94
REFINED AUC 0.76

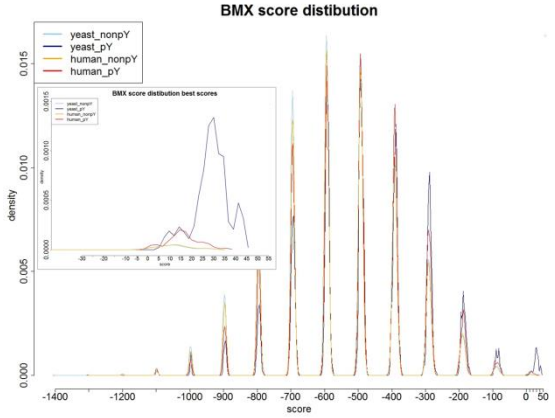

TNK1 motif
True positive rate
False positive rate
ORIGINAL AUC 0.82
REFINED AUC 0.64

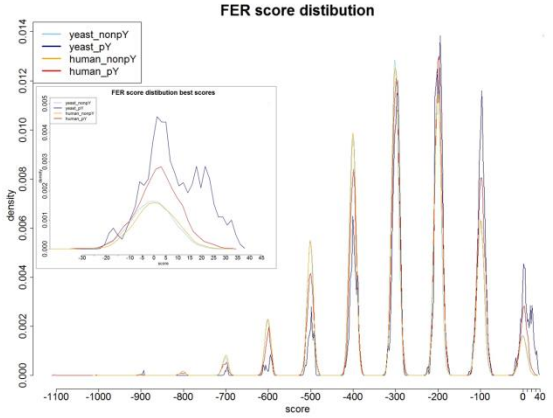

YES1 motif
True positive rate
False positive rate
ORIGINAL AUC 0.89
REFINED AUC 0.79

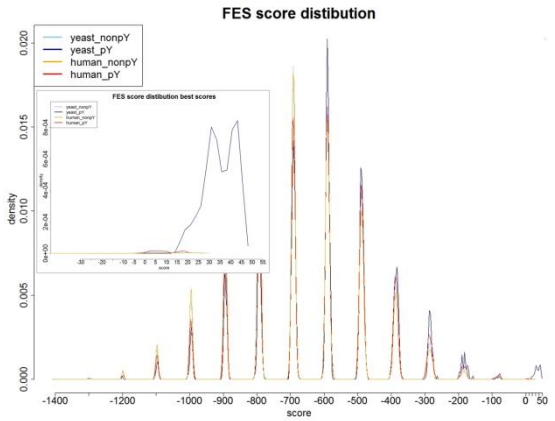
6.5. Motif score distributions for reported human pY-sites

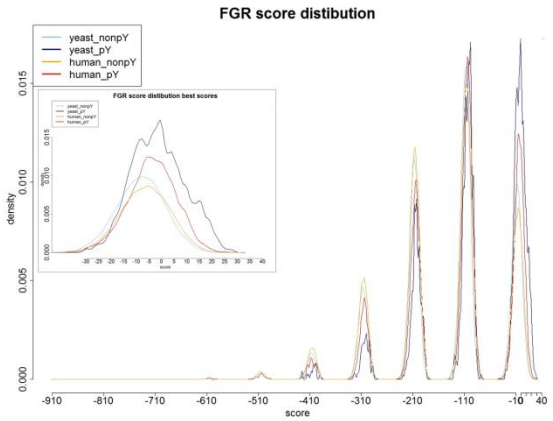


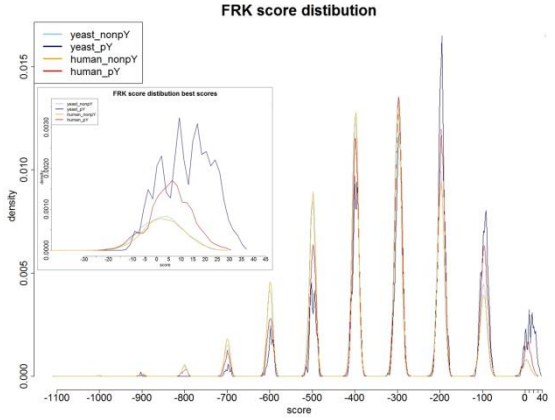


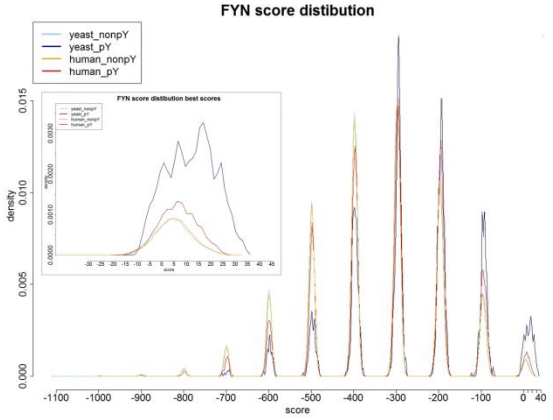


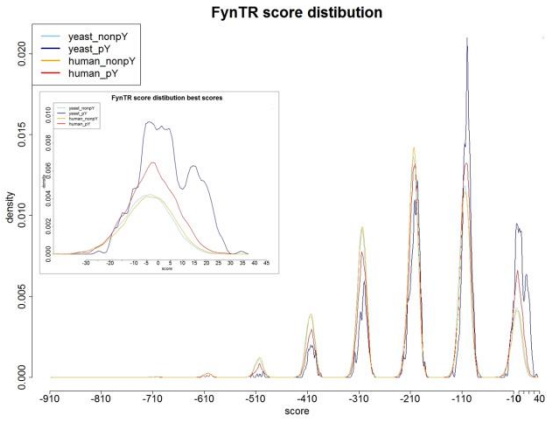


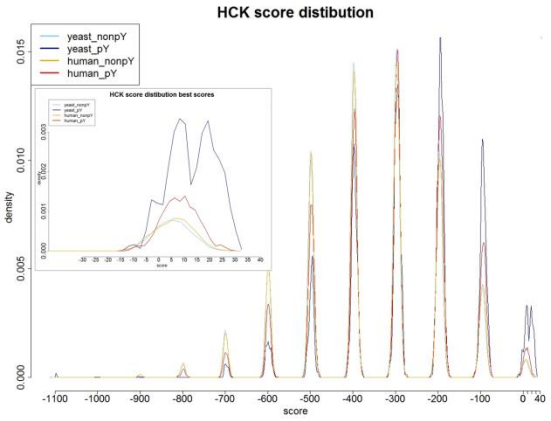


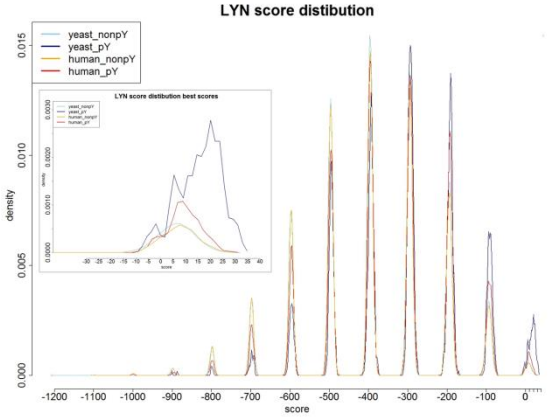


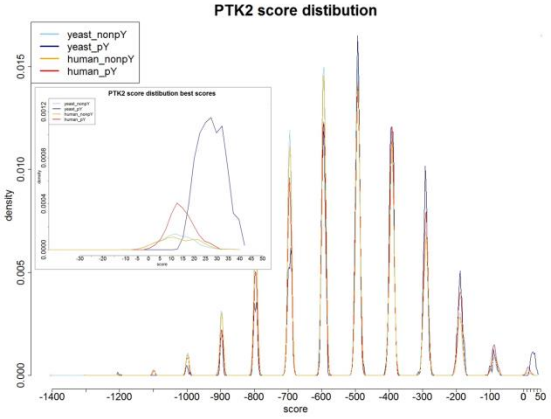


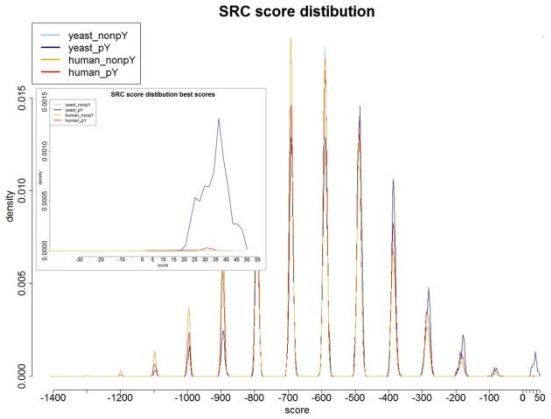


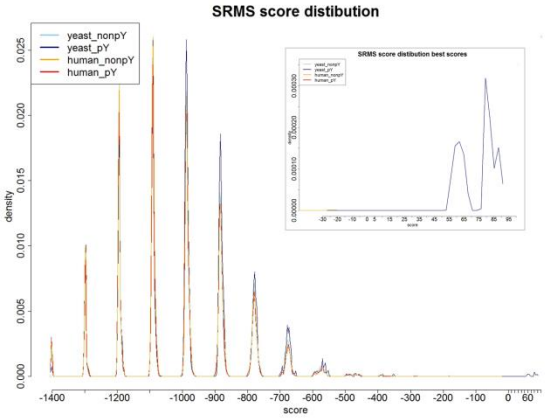

FynTR score distibution
yeast_nonpY
yeast_pY
human_nonpY
human_pY
density
score

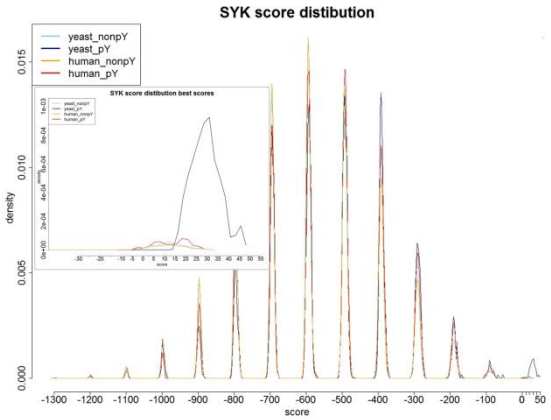

HCK score distibution
yeast_nonpY
yeast_pY
human_nonpY
human_pY
density
score

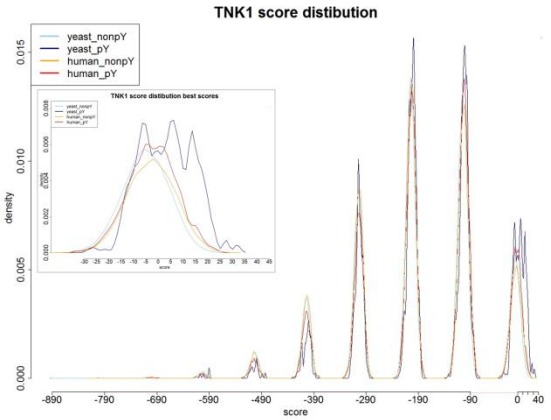

LYN score distibution
yeast_nonpY
yeast_pY
human_nonpY
human_pY
density
score

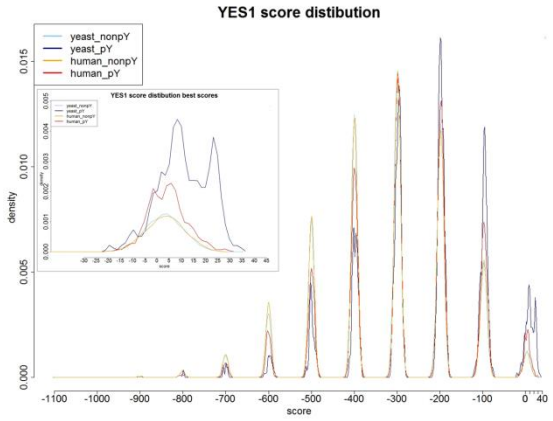

PTK2 score distibution
yeast_nonpY
yeast_pY
human_nonpY
human_pY
density
score


SRC score distibution
yeast_nonpY
yeast_pY
human_nonpY
human_pY
density
score


SRMS score distibution
yeast_nonpY
yeast_pY
human_nonpY
human_pY
density
score


SYK score distibution
yeast_nonpY
yeast_pY
human_nonpY
human_pY
density
score


TNK1 score distibution
yeast_nonpY
yeast_pY
human_nonpY
human_pY
density
score



6.6. Motif predicted human kinase-substrate relationships

Symbol	GeneID	RefSeqID	Site	Sequence	Predicted Kinases
A1CF	29974	NP_620310	Y404	GGRGYLAYTGLGRGY	FER, FGR, TNK1
AARS	16	NP_001596	Y580	VLHIGTIYGDLKVGD	BLK
ABCA3	21	NP_001080	Y1349	RRTLTELYTRMPVLP	FRK
ABI1	10006	NP_005461	Y455	IEKVVAIYDYTKDKD	FynTR, SRC, SYK, YES1
ABLIM1	3983	NP_002304	Y357	RTSSESIYSRPGSSI	ABL2, FER, FRK
ABLIM1	3983	NP_002304	Y396	IPKVKAIYDIERPDL	FYN
ABLIM3	22885	NP_055760	Y361	SPYSQDIYENLDLRQ	YES1
ACE2	59272	NP_068576	Y781	ARSGENPYASIDISK	FGR, FRK, SRC, YES1
ACIN1	22985	NP_055792	Y512	QKSTLADYSAQKDLE	BMX
ACLY	47	NP_001087	Y682	SRTTDGVYEGVAIGG	FGR, FRK, HCK, YES1
ACO1	48	NP_002188	Y466	AGLNVMPYIKTSLSP	PTK2
ACOX1	51	NP_004026	Y275	PLSNKLTYGTMVFVR	HCK
ACOX2	8309	NP_003491	Y417	RACGGHGYSKLSGLP	TNK1
ACSBG1	23205	NP_055977	Y138	EHISYSQYYLLARRA	FRK
ACTA1	58	NP_001091	Y55	GMGQKDSYVGDEAQS	PTK2
ACTA2	59	NP_001604	Y55	GMGQKDSYVGDEAQS	PTK2
ACTB	60	NP_001092	Y53	GMGQKDSYVGDEAQS	PTK2
ACTB	60	NP_001092	Y166	VTHTVPIYEGYALPH	BLK, FER, FYN, SRC, YES1
ACTC1	70	NP_005150	Y55	GMGQKDSYVGDEAQS	PTK2
ACTG1	71	NP_001605	Y53	GMGQKDSYVGDEAQS	PTK2
ACTG1	71	NP_001605	Y166	VTHTVPIYEGYALPH	BLK, FER, FYN, SRC, YES1
ACTG2	72	NP_001606	Y54	GMGQKDSYVGDEAQS	PTK2
ACTR10	55860	NP_060947	Y4	----MPLYEGLGSGG	LYN
ACTR3	10096	NP_005712	Y16	VVDCGTGYTKLGYAG	LYN
AFAP1L2	84632	NP_001001936	Y457	TDPEEFTYDYVDADR	SYK
AGAP2	116986	NP_001116244	Y1130	QGRTALFYARQAGSQ	PTK2
AGAP3	116988	NP_114152	Y861	RGLTPLAYARRAGSQ	FGR, PTK2, TNK1
AHCTF1	25909	NP_056261	Y1799	SEASENIYSDVRGLS	FER, FES
AHNAK	79026	NP_001611	Y160	VTRRVTAYTVDVTGR	TNK1
AIP	9049	NP_003968	Y203	KEAAAKYYDAIACLK	FER, FES, HCK

AIP	9049	NP_003968	Y248	KLVVEEYYEVLDHCS	FynTR
AIPL1	23746	NP_055151	Y247	LLKKEEYYEVLEHTS	BLK, HCK
AK7	122481	NP_689540	Y390	IAKELANYYKLHHIQ	SYK
AKAP11	11215	NP_057332	Y488	HDSVYYTYEDYAKSI	FER
AKAP12	9590	NP_005091	Y374	EPRLSAEYEKVELPS	FER
ALDH1A3	220	NP_000684	Y437	KRANSTDYGLTAAVF	FGR
ALDOA	226	NP_000025	Y3	-----MPYQYPALTP	PTK2, SRMS
ALDOA	226	NP_000025	Y364	LFVSNHAY-------	SRMS
ALKBH3	221120	NP_631917	Y127	GIREDITYQQPRLTA	ABL2, FRK
ALPP	250	NP_001623	Y415	KAYTVLLYGNGPGYV	TNK1
ALPPL2	251	NP_112603	Y412	KAYTVLLYGNGPGYV	TNK1
AMBRA1	55626	NP_060219	Y625	GSREHPIYPDPARLS	FES
AMBRA1	55626	NP_060219	Y636	ARLSPAAYYAQRMIQ	PTK2
AMOTL1	154810	NP_570899	Y191	NNEELPTYEEAKAQS	SYK
AMOTL2	51421	NP_057285	Y107	KGEELPTYEEAKAHS	SYK
ANKIB1	54467	NP_061877	Y216	AEEIEAEYAALDKRE	BMX
ANKRD11	29123	NP_037407	Y214	EACNRGYYDVAKQLL	BMX, FER, FYN
ANKRD12	23253	NP_056023	Y943	KQSDNSEYSKSEKGK	ABL2
ANKRD26	22852	NP_055730	Y296	RKNLEATYGTVRTGN	FER, FynTR, YES1
ANKS1A	23294	NP_056060	Y834	ASLADRPYEEPPQKP	FGR
ANTXR1	84168	NP_115584	Y383	PTVDASYYGGRGVGG	HCK
ANTXR2	118429	NP_001139266	Y381	PTVDASYYGGRGVGG	HCK
ANXA2	302	NP_001002858	Y334	RKYGKSLYYYIQQDT	FER, HCK
ANXA2	302	NP_001002858	Y351	DYQKALLYLCGGDD-	SYK
ANXA3	306	NP_005130	Y300	KHYGYSLYSAIKSDT	ABL2, FRK, FYN, FynTR, PTK2
ANXA5	308	NP_001145	Y257	YLAETLYYAMKGAGT	BLK
ANXA6	309	NP_001146	Y302	TKYEKSLYSMIKNDT	SYK
AP1AR	55435	NP_061039	Y110	RLEEEALYAAQREAA	FRK
AP1B1	162	NP_001118	Y136	CLKDEDPYVRKTAAV	FRK
AP2B1	163	NP_001025177	Y136	CLKDEDPYVRKTAAV	FRK
APC	324	NP_001120983	Y2645	AESKTLIYQMAPAVS	BMX, FynTR, SYK
APEX1	328	NP_542380	Y45	AGEGPALYEDPPDQK	FRK
ARAP1	116985	NP_001035207	Y343	KNPPQGSYIYQKRWV	FRK
ARAP3	64411	NP_071926	Y247	EAREDAGYASLELPG	FRK, SRC
ARAP3	64411	NP_071926	Y1408	PVYEEPVYEEVGAFP	FRK, FYN, FynTR
ARF1	375	NP_001649	Y167	ATSGDGLYEGLDWLS	FRK, FynTR, YES1
ARF3	377	NP_001650	Y167	ATSGDGLYEGLDWLA	FRK, FynTR, YES1
ARF6	382	NP_001654	Y163	ATSGDGLYEGLTWLT	FRK, FynTR, YES1
ARFGAP1	55738	NP_783202	Y208	NNAMSSLYSGWSSFT	YES1
ARFGAP2	84364	NP_115765	Y445	GREVDAEYEARSRLQ	BMX, FRK, SYK
ARFGAP3	26286	NP_055385	Y441	GRQSQADYETRARLE	SYK
ARHGAP30	257106	NP_001020769	Y1027	TCTEGGDYCLIPRTS	FGR
ARHGAP32	9743	NP_001136157	Y1702	RLQGKSLYSYAGLAP	FER
ARHGAP32	9743	NP_001136157	Y1704	QGKSLYSYAGLAPRP	TNK1
ARHGAP32	9743	NP_001136157	Y1788	QDDLGGIYVIHLRSK	FYN
ARHGAP32	9743	NP_001136157	Y2005	ERDPSVLYQYQPHGK	FGR
ARHGAP33	115703	NP_443180	Y1008	PTPPEPLYVNLALGP	FES, YES1
ARHGDIA	396	NP_004300	Y156	YGPRAEEYEFLTPVE	FES
ARHGEF1	9138	NP_945353	Y753	WDQEAQIYELVAQTV	SRC
ARHGEF15	22899	NP_776089	Y353	PLQDEPLYQTYRAAV	ABL2
ARHGEF2	9181	NP_001155855	Y434	SNVDEGIYQLEKGAR	FynTR

ARHGEF26	26084	NP_056410	Y380	GEENAVLYQNYKEKA	LYN, SYK
ARHGEF5	7984	NP_005426	Y641	RPPKPAIYSSVTPRR	SRC
ARHGEF5	7984	NP_005426	Y1097	INSSQLLYQEYSDVV	FRK
ARHGEF7	8874	NP_001106983	Y620	PAPLTPAYHTLPHPS	ABL2
ARID1B	57492	NP_059989	Y1362	KRHMDGMYGPPAKRH	BLK
ARL11	115761	NP_612459	Y30	AGKTTLLYKLKGHQL	BMX, LYN
ASAP3	55616	NP_060177	Y733	LDISNKTYETVASLG	LYN
ASB4	51666	NP_057200	Y426	LEPEGIIY-------	FES, SRMS
ATAD2	29028	NP_054828	Y322	QRKPNIFYSGPASPA	FRK
ATG14	22863	NP_055739	Y37	VDDAEGLYVAVERCP	SRC
ATIC	471	NP_004035	Y303	LTPISAAYARARGAD	TNK1
ATP10A	57194	NP_077816	Y258	NGKKAGLYKENLLLR	BMX
ATP13A1	57130	NP_065143	Y202	PVGNAFSYYQSNRGF	TNK1
ATP2A1	487	NP_775293	Y497	DRKSMSVYCSPAKSS	LYN
ATP2B2	491	NP_001001331	Y1152	KAFRSSLYEGLEKPE	BLK
ATP5B	506	NP_001677	Y361	ITSVQAIYVPADDLT	BMX
ATP5O	539	NP_001688	Y35	VRPPVQVYGIEGRYA	SRC
ATP5O	539	NP_001688	Y46	GRYATALYSAASKQN	FRK, FYN
ATP6V0A2	23545	NP_036595	Y149	NVEFEPTYEEFPSLE	HCK
ATP6V1D	51382	NP_057078	Y119	YHEGTDSYELTGLAR	FRK
ATP6V1H	51606	NP_998785	Y388	VRLNEKNYELLKILT	FER, FES
ATP7A	538	NP_000043	Y517	LRREEGIYSILVALM	BMX, FES, FYN, HCK
ATP9B	374868	NP_940933	Y25	ANRKRAAYYSAAGPR	FYN
ATP9B	374868	NP_940933	Y69	FENEESDYHTLPRAR	ABL2
ATPBD4	89978	NP_542381	Y97	GDEVEDLYELLKLVK	BLK, BMX, FRK, FYN, FynTR, HCK
AXIN2	8313	NP_004646	Y460	QSPGVGRYSPRSRSP	SRC
AXL	558	NP_068713	Y759	GVENSEIYDYLRQGN	BLK, FES, FRK, FYN, FynTR
B3GAT1	27087	NP_473366	Y282	LRGVKGGYQESSLLR	SRMS
BAG3	9531	NP_004272	Y240	YPAQQGEYQTHQPVY	BMX
BAK1	578	NP_001179	Y110	TAENAYEYFTKIATS	SYK
BCAR1	9564	NP_055382	Y128	SKAQQGLYQVPGPSP	ABL2, FER, HCK, PTK2
BCAR1	9564	NP_055382	Y179	GGPAQDIYQVPPSAG	BLK, FRK, LYN
BCAR1	9564	NP_055382	Y234	AQPEQDEYDIPRHLL	ABL2
BCAR1	9564	NP_055382	Y249	APGPQDIYDVPPVRG	YES1
BCAR1	9564	NP_055382	Y287	RDPLLEVYDVPPSVE	BLK
BCAR1	9564	NP_055382	Y327	PLLREETYDVPPAFA	BLK, FynTR
BCAR1	9564	NP_055382	Y362	SPPAEDVYDVPPPAP	ABL2
BCAR1	9564	NP_055382	Y387	RPGPGTLYDVPRERV	LYN
BCAR1	9564	NP_055382	Y410	GVVDSGVYAVPPPAE	FER
BCKDHB	594	NP_000047	Y392	ALRKMINY-------	FES
BCLAF1	9774	NP_055554	Y80	YRGRGRGYYQGGGGR	TNK1
BCLAF1	9774	NP_055554	Y81	RGRGRGYYQGGGGRY	PTK2, TNK1
BCLAF1	9774	NP_055554	Y219	IWPGLSAYDNSPRSP	PTK2
BCOR	54880	NP_001116857	Y972	RIANSAGYVGDRFKC	PTK2
BCR	613	NP_004318	Y279	PPLEYQPYQSIYVGG	HCK
BET1	10282	NP_005859	Y18	PPGNYGNYGYANSGY	TNK1
BET1	10282	NP_005859	Y20	GNYGNYGYANSGYSA	PTK2
BIRC6	57448	NP_057336	Y4130	EQSGELVYEAPETVA	BLK, FER, FynTR
BLMH	642	NP_000377	Y420	EWFSEYVYEVVVDRK	YES1

BMP2K	55589	NP_942595	Y1012	KKTLKPTYRTPERAR	ABL2, LYN
BPTF	2186	NP_872579	Y1742	GGIREVPYFNYNAKP	PTK2
BRAF	673	NP_004324	Y746	QTEDFSLYACASPKT	ABL2
BRD7	29117	NP_037395	Y217	YNKPETIYYKAAKKL	FRK, FYN, LYN
BRWD1	54014	NP_061836	Y19	PLIESELYFLIARYL	HCK
BSN	8927	NP_003449	Y2068	YSSVSNIYSDHRYGP	FynTR
BTBD10	84280	NP_115696	Y377	GIEGYPTYKEKVKKR	HCK, LYN
BTNL8	79908	NP_001035552	Y112	TVLDAGLYGCRISSQ	FYN
BUD13	84811	NP_116114	Y280	DLAPNVTYSLPRTKS	FER, LYN
BUD13	84811	NP_116114	Y494	DSERDELYAQWGKGL	BLK
C10orf54	64115	NP_071436	Y265	KVRHPLSYVAQRQPS	PTK2
C11orf52	91894	NP_542390	Y78	SEDSNLHYADIQVCS	FER
C17orf39	79018	NP_076957	Y297	PEHSAPIYEFR----	BMX, HCK, SRC
C18orf34	374864	NP_001098998	Y422	LEAVNDFYAAKKTWD	FRK
C19orf25	148223	NP_689695	Y63	EAPGEQLYQQSRAYV	ABL2
C1orf150	148823	NP_660321	Y69	SGSEEVCYTVINHIP	BMX
C1orf150	148823	NP_660321	Y110	RERSETEYALLRTSV	BLK, FRK
C3orf24	115795	NP_775743	Y163	LRSILLLYATYKKCT	BLK, BMX
C5orf24	134553	NP_689622	Y35	AADQFDIYSSQQSKY	YES1
C5orf24	134553	NP_689622	Y134	GTTKAAGYKVSPGRP	PTK2, TNK1
C6orf25	80739	NP_612116	Y237	PADASTIYAVVV---	FER, FYN, LYN, YES1
C9orf78	51759	NP_057604	Y147	KNAEDCLYELPENIR	LYN
C9orf78	51759	NP_057604	Y279	KATDDYHYEKFKKMN	SRC
CAMK2B	816	NP_001211	Y17	FTDEYQLYEDIGKGA	FER
CAMKK2	10645	NP_006540	Y128	CICPSLPYSPVSSPQ	FER
CAMKK2	10645	NP_006540	Y190	YNENDNTYYAMKVLS	FRK
CAP1	10487	NP_006358	Y354	LKQVAYIYKCVNTTL	SRMS
CAPN7	23473	NP_055111	Y619	KTDGKKVYYPADPPP	FER
CAPRIN1	4076	NP_005889	Y283	EPEPAEEYTEQSEVE	BMX
CAPRIN1	4076	NP_005889	Y541	TLKQQNQYQASYNQS	ABL2
CAPZA3	93661	NP_201585	Y289	PKLGYVIYSRSVLCN	ABL2, FER
CASC5	57082	NP_733468	Y443	AMTPESIYSNPSIQG	PTK2
CASP8	841	NP_001073594	Y393	DGQEAPIYELTSQFT	FES
CASP8	841	NP_001073594	Y439	TDSEEQPYLEMDLSS	SRC
CASS4	57091	NP_065089	Y98	PASSEETYQVPTLPR	ABL2, BLK, FER, FRK
CASS4	57091	NP_065089	Y195	EPEKQQLYDIPASPK	ABL2, BLK, HCK
CASS4	57091	NP_065089	Y231	TTLRRGGYSTLPNPQ	LYN
CAT	847	NP_001743	Y231	NANGEAVYCKFHYKT	FynTR, HCK, SRC
CAT	847	NP_001743	Y386	YRARVANYQRDGPMC	PTK2
CBL	867	NP_005179	Y674	SSSANAIYSLAARPL	FER
CBLB	868	NP_733762	Y665	GHLGSEEYDVPPRLS	ABL2
CC2D1A	54862	NP_060191	Y207	GPASTPTYSPAPTQP	FER, LYN
CCDC120	90060	NP_296375	Y398	CKSSEVLYERPQPTP	FER
CCDC50	152137	NP_848018	Y144	TRAYADSYYYEDGDQ	SYK
CCDC50	152137	NP_848018	Y145	RAYADSYYYEDGDQP	SYK
CCDC50	152137	NP_848018	Y480	SSHKGFHYKH-----	SRMS
CCDC59	29080	NP_054886	Y176	NQKAQEEYEQIQAKR	ABL2, BLK
CCDC88A	55704	NP_001129069	Y1798	QSKDSNPYATLPRAS	FGR, TNK1
CCNB1IP1	57820	NP_067001	Y178	YQKLQGLYDSLRLRN	BLK
CCT8	10694	NP_006576	Y30	SGLEEAVYRNIQACK	BMX
CD200R1	131450	NP_620161	Y325	TEKNNPLYDTTNKVK	BMX, FES
CD22	933	NP_001762	Y842	QAQENVDYVILKH--	ABL2

CD226	10666	NP_006557	Y322	DDTREDIYVNYPTFS	PTK2
CD2AP	23607	NP_036252	Y541	ELKKDTCYSPKPSVY	BLK
CD300A	11314	NP_009192	Y255	PREVEVEYSTVASPR	ABL2, BLK, FGR, SYK, YES1
CD300LF	146722	NP_620587	Y284	GPEEPTEYSTISRP-	ABL2
CD34	947	NP_001020280	Y329	ERLGEDPYYTENGGG	FGR
CD3D	915	NP_000723	Y149	LLRNDQVYQPLRDRD	BLK
CD3E	916	NP_000724	Y188	PPVPNPDYEPIRKGQ	FER
CD3E	916	NP_000724	Y199	RKGQRDLYSGLNQRR	FGR
CD3G	917	NP_000064	Y160	LLPNDQLYQPLKDRE	BLK, HCK, LYN
CD46	4179	NP_758869	Y384	KADGGAEYATYQTKS	FER, YES1
CD46	4179	NP_758869	Y387	GGAEYATYQTKSTTP	FRK, PTK2
CD5	921	NP_055022	Y487	DNSSDSDYDLHGAQR	FYN
CD59	966	NP_000602	Y86	LRENELTYYCCKKDL	FYN, HCK
CD7	924	NP_006128	Y239	TLSSPNQYQ------	PTK2
CD79A	973	NP_001774	Y188	EYEDENLYEGLNLDD	FRK
CD79A	973	NP_001774	Y210	SRGLQGTYQDVGSLN	FES
CD79B	974	NP_001035022	Y197	GMEEDHTYEGLDIDQ	FYN
CD79B	974	NP_001035022	Y208	DIDQTATYEDIVTLR	FER
CD82	3732	NP_002222	Y267	DYSKVPKY-------	SRC, SRMS
CD84	8832	NP_003865	Y299	EEPVNTVYSEVQFAD	BLK
CD8B	926	NP_004922	Y209	LRFMKQFYK------	SRMS
CDC27	996	NP_001107563	Y746	VPKESLVYFLIGKVY	FER, FynTR, HCK, SRC
CDC42EP4	23580	NP_036253	Y255	TITQAPPYAVAAPPL	TNK1
CDC5L	988	NP_001244	Y459	PEDGMADYSDPSYVK	FER
CDC73	79577	NP_078805	Y290	KQPIPAAYNRYDQER	TNK1
CDH2	1000	NP_001783	Y785	GGEEDQDYDLSQLQQ	BMX
CDH2	1000	NP_001783	Y860	DSLLVFDYEGSGSTA	SRC
CDH23	64072	NP_071407	Y3195	LRAAIQEYDNIAKLG	FES
CDH4	1002	NP_001785	Y795	GGEEDQDYDLSQLQQ	BMX
CDH7	1005	NP_004352	Y118	DREEQAYYTLRAQAL	FYN
CDK1	983	NP_001777	Y15	EKIGEGTYGVVYKGR	BLK, FER, FGR, FYN, FynTR, HCK
CDK11A	728642	NP_076916	Y434	NRIEEGTYGVVYRAK	BLK, FYN, LYN
CDK11B	984	NP_277021	Y436	NRIEEGTYGVVYRAK	BLK, FYN, LYN
CDK13	8621	NP_003709	Y362	LPRSPSPYSRRRSPS	HCK
CDK13	8621	NP_003709	Y716	GIIGEGTYGQVYKAR	BLK
CDK14	5218	NP_036527	Y128	EKLGEGSYATVYKGK	FER, SRC
CDK15	65061	NP_631897	Y63	EKLGEGSYATVYKGI	FER
CDK16	5127	NP_148978	Y182	DKLGEGTYATVYKGK	BLK, FER, FGR, FYN, FynTR, HCK, LYN, SRC, YES1
CDK17	5128	NP_002586	Y203	EKLGEGTYATVYKGR	BLK, FER, FGR, FRK, FYN, FynTR, HCK, SRC, YES1
CDK18	5129	NP_997668	Y185	DKLGEGTYATVFKGR	BLK, FGR, FYN, FynTR, HCK, SRC
CDK2	1017	NP_001789	Y15	EKIGEGTYGVVYKAR	BLK, FER, FYN
CDK3	1018	NP_001249	Y15	EKIGEGTYGVVYKAK	BLK, FER, FYN
CDK5	1020	NP_004926	Y15	EKIGEGTYGTVFKAK	BLK, FYN, SRC
CDKL5	6792	NP_003150	Y171	NNANYTEYVATRWYR	FGR, FRK

CDKN1B	1027	NP_004055	Y88	KGSLPEFYYRPPRPP	FER, FRK
CEACAM1	634	NP_001703	Y520	LTATEIIYSEVKKQ-	BMX, FYN
CEBPE	1053	NP_001796	Y107	KALGPGIYSSPGSYD	FynTR, SRC, YES1
CECR6	27439	NP_114096	Y534	PSRARGGYGAPPSAP	LYN, TNK1
CELSR3	1951	NP_001398	Y3047	RAVPAASYGRIYAGG	TNK1
CENPF	1063	NP_057427	Y1731	EQSPDTNYEPPGEDK	FER
CENPF	1063	NP_057427	Y3000	PTGKTSPYILRRTTM	FRK
CEP350	9857	NP_055625	Y2611	REKDVSEYFYEKSLP	SYK
CEP89	84902	NP_116205	Y157	GGHSDDLYAVPHRNQ	HCK
CFL1	1072	NP_005498	Y68	GQTVDDPYATFVKML	FynTR
CFL1	1072	NP_005498	Y89	YALYDATYETKESKK	SRC
CFL2	1073	NP_068733	Y89	YALYDATYETKESKK	SRC
CGN	57530	NP_065821	Y105	LELPENPYSQVKGFP	FGR
CHERP	10523	NP_006378	Y714	GWEQNGLYEFFRAKM	FynTR, HCK
CHGB	1114	NP_001810	Y526	GELFNPYYDPLQWKS	FER
CHI3L2	1117	NP_003991	Y82	DKSEVMLYQTINSLK	ABL2, SRC
CHKA	1119	NP_001268	Y197	RSLGPKLYGIFPQGR	FGR, HCK, SRC
CHMP6	79643	NP_078867	Y136	ETQEAVEYQRQIDEL	SYK
CHN2	1124	NP_004058	Y153	KMTTNPIYEHIGYAT	FynTR
CHRM3	1131	NP_000731	Y365	GSETRAIYSIVLKLP	BMX
CHST15	51363	NP_056976	Y221	LTNSYVLYSKRFRST	BLK
CIRBP	1153	NP_001271	Y142	YGGSRDYYSSRSQSG	LYN, PTK2
CIT	11113	NP_009105	Y1478	EGSKVLIYDNEAREA	SYK
CKB	1152	NP_001814	Y39	KVLTPELYAELRAKS	BLK, HCK
CLASP1	23332	NP_056097	Y1179	NLNSEEIYSSLRGVT	BLK, FRK, FYN, FynTR, SRC
CLASP2	23122	NP_055912	Y1150	NMNSEDIYSSLRGVT	FYN, FynTR
CLCC1	23155	NP_001041675	Y531	SPDQGSTYSPARGVA	FER
CLCN5	1184	NP_001121370	Y495	RPAGVGVYSAMWQLA	SRC
CLDN3	1365	NP_001297	Y198	YTATKVVYSAPRSTG	ABL2, BLK
CLDN3	1365	NP_001297	Y219	TGYDRKDYV------	SYK
CLDN4	1364	NP_001296	Y208	RSAAASNYV------	PTK2
CLDN6	9074	NP_067018	Y219	SEYPTKNYV------	FES
CLDN7	1366	NP_001298	Y210	KSNSSKEYV------	FES
CLDN9	9080	NP_066192	Y200	PRGPRLGYSIPSRSG	LYN
CLDND1	56650	NP_001035272	Y267	AHTNRKEYTLMKAYR	FRK
CLIC1	1192	NP_001279	Y233	DEEIELAYEQVAKAL	BLK
CLIC2	1193	NP_001280	Y239	DKEIENTYANVAKQK	LYN, YES1
CLINT1	9685	NP_055481	Y293	DLGAAAHYTGDKASP	PTK2
CLK2	1196	NP_003984	Y51	RRRREDSYHVRSRSS	PTK2
CLTA	1211	NP_009027	Y94	SNGPTDSYAAISQVD	FRK
CLTC	1213	NP_004850	Y634	LEHFTDLYDIKRAVV	BLK, HCK
CMYA5	202333	NP_705838	Y209	PPITGAIYKEHKPLV	BMX, SRMS
CNKSR3	154043	NP_775786	Y365	DENGSFVYGGSSKCK	SRC
CNN1	1264	NP_001290	Y261	SQRGMTVYGLPRQVY	FER, LYN
CNN3	1266	NP_001830	Y261	SQKGMSVYGLGRQVY	LYN
COBL	23242	NP_056013	Y652	AKVKDKVYGCADGER	FYN
COBL	23242	NP_056013	Y867	QRRTSSQYVASAIAK	ABL2
COL14A1	7373	NP_066933	Y988	SEYKISVYTKLQEIE	BMX
COL17A1	1308	NP_000485	Y40	KGGTSNGYAKTASLG	LYN
COL4A4	1286	NP_000083	Y1541	AQRNDRSYWLASAAP	TNK1
CORO1A	11151	NP_009005	Y25	PAKADQCYEDVRVSQ	BLK

CORO7	79585	NP_078811	Y288	GERQLYCYEVVPQQP	BMX
CORO7	79585	NP_078811	Y758	GDTRVFLYELLPESP	FynTR, SRC
CPD	1362	NP_001295	Y1376	DTEEETLYSSKH---	HCK, SRC
CPNE4	131034	NP_570720	Y75	RTCINPVYSKLFTVD	BLK
CPQ	10404	NP_057218	Y56	AIINLAVYGKAQNRS	BMX
CPSF6	11052	NP_008938	Y384	GPPPTDPYGRPPPYD	TNK1
CRAT	1384	NP_000746	Y110	LKTAYLQYRQPVVIY	ABL2
CRCP	27297	NP_055293	Y47	QNLNTITYETLKYIS	YES1
CRIP1	1396	NP_001302	Y12	PKCNKEVYFAERVTS	HCK
CRK	1398	NP_058431	Y221	GGPEPGPYAQPSVNT	ABL2, TNK1
CRK	1398	NP_058431	Y239	NLQNGPIYARVIQKR	ABL2
CRKL	1399	NP_005198	Y207	IPEPAHAYAQPQTTT	TNK1
CRMP1	1400	NP_001014809	Y618	GMYDGPVYEVPATPK	FynTR
CRTC1	23373	NP_001091952	Y61	GPSRGQYYGGSLPNV	TNK1
CSDE1	7812	NP_001123995	Y612	EEADPTIYSGKVIRP	HCK
CSF2RB	1439	NP_000386	Y628	PPPGSLEYLCLPAGG	ABL2
CSF3R	1441	NP_724781	Y754	LPTLVQTYVLQGDPR	YES1
CSNK2A1	1457	NP_001886	Y255	VLGTEDLYDYIDKYN	FYN
CSNK2A2	1459	NP_001887	Y256	VLGTEELYGYLKKYH	BLK, FYN
CSRP1	1465	NP_004069	Y73	YGPKGYGYGQGAGTL	SRMS, TNK1
CSRP1	1465	NP_004069	Y127	PRCSQAVYAAEKVIG	FYN, HCK
CST1	1469	NP_001889	Y64	KATKDDYYRRPLRVL	ABL2
CSTB	1476	NP_000091	Y97	AKHDELTYF------	SRC
CTLA4	1493	NP_005205	Y201	SPLTTGVYVKMPPTE	YES1
CTNNA1	1495	NP_001894	Y619	IDASRLVYDGIRDIR	LYN
CTNNAL1	8727	NP_003789	Y436	NLEALAEYACKLSEQ	SYK
CTNNB1	1499	NP_001895	Y654	RNEGVATYAAAVLFR	FYN
CTNND1	1500	NP_001078927	Y193	GNGGPGPYVGQAGTA	PTK2
CTNND2	1501	NP_001323	Y292	SAPEGATYAAPRGSS	FER, PTK2
CTNND2	1501	NP_001323	Y499	AAGPASNYADPYRQL	PTK2
CTNND2	1501	NP_001323	Y516	CPSVESPYSKSGPAL	FER
CTSO	1519	NP_001325	Y145	VGAVESAYAIKGKPL	BLK
CTTN	2017	NP_005222	Y302	KHESQQDYSKGFGGK	FYN
CTTN	2017	NP_005222	Y334	VTQVSSAYQKTVPVE	FER
CTTN	2017	NP_005222	Y421	RLPSSPVYEDAASFK	FES
CTTN	2017	NP_005222	Y470	AYATEAVYESAEAPG	FRK
CTTNBP2NL	55917	NP_061174	Y447	SSASSPGYQSSYQVG	PTK2
CYTH1	9267	NP_004753	Y382	AISRDPFYEMLAARK	FynTR
DAB1	1600	NP_066566	Y220	PETEENIYQVPTSQK	ABL2, FER, FRK, FynTR
DAB1	1600	NP_066566	Y232	SQKKEGVYDVPKSQP	ABL2
DAB2	1601	NP_001334	Y685	TLSAFASYFNSKVGI	TNK1
DAG1	1605	NP_004384	Y863	EDPNAPPYQPPPPFT	PTK2
DAG1	1605	NP_004384	Y892	PYRSPPPYVPP----	PTK2
DAPP1	27071	NP_055210	Y139	KVEEPSIYESVRVHT	HCK
DAXX	1616	NP_001135442	Y136	RSRPAKLYVYINELC	FES
DBNL	28988	NP_001116428	Y162	QAPVGSVYQKTNAVS	BMX
DBNL	28988	NP_001116428	Y343	QAEEEAVYEEPPEQE	BMX
DBR1	51163	NP_057300	Y533	KRRNQAIYAAVDDDD	FRK
DCBLD2	131566	NP_563615	Y565	KKKTEGTYDLPYWDR	FER, FGR, FRK
DCBLD2	131566	NP_563615	Y655	GYADLDPYNSPGQEV	FGR
DCBLD2	131566	NP_563615	Y750	PAPDELVYQVPQSTQ	ABL2, BLK, FER,

					FynTR, HCK
DCC	1630	NP_005206	Y1363	AIEPKVPYTPLLSQP	SRMS
DDR1	780	NP_054700	Y484	GPREPPPYQEPRPRG	ABL2, HCK
DDR1	780	NP_054700	Y802	RNLYAGDYYRVQGRA	FGR
DDR2	4921	NP_006173	Y740	RNLYSGDYYRIQGRA	FGR
DDX17	10521	NP_001091974	Y279	RLKSTCIYGGAPKGP	FER, FGR, FYN, FynTR, LYN
DDX26B	203522	NP_872346	Y513	PQTYRNAYDIPRRGL	FER
DDX3X	1654	NP_001347	Y104	DDRGRSDYDGIGSRG	FGR
DDX3X	1654	NP_001347	Y283	RELAVQIYEEARKFS	FRK
DDX3Y	8653	NP_004651	Y258	AVKENGRYGRRKQYP	TNK1
DDX3Y	8653	NP_004651	Y281	RELAVQIYEEARKFS	FRK
DDX41	51428	NP_057306	Y33	EDEDDEDYVPYVPLR	BMX
DDX5	1655	NP_004387	Y202	RLKSTCIYGGAPKGP	FER, FGR, FYN, FynTR, LYN
DDX5	1655	NP_004387	Y593	NGMNQQAYAYPATAA	BLK
DHX36	170506	NP_065916	Y1007	PRFQDGYYS------	SRMS
DHX9	1660	NP_001348	Y9	GDVKNFLYAWCGKRK	FER, FynTR
DHX9	1660	NP_001348	Y1241	YRGPSGGYRGSGGFQ	PTK2
DIP2C	22982	NP_055789	Y213	APPDVTTYTSEHSIQ	HCK
DLG3	1741	NP_066943	Y306	KNTSDMVYLKVAKPG	SRC
DLG4	1742	NP_001356	Y279	VAALKNTYDVVYLKV	BLK, SRC
DLG4	1742	NP_001356	Y744	GDSFEEIYHKVKRVI	ABL2
DLG5	9231	NP_004738	Y429	REERDAVYSEYKLIM	BMX
DLG5	9231	NP_004738	Y1197	SILRNPIYTVRSHRV	BMX
DLGAP2	9228	NP_004736	Y229	NLDSDSTYRTPSVLN	ABL2
DLX2	1746	NP_004396	Y122	AYTSYAPYGTSSSPA	PTK2, TNK1
DNAJB1	3337	NP_006136	Y176	RVSLEEIYSGCTKKM	FynTR
DNAJB4	11080	NP_008965	Y172	RVSLEEIYSGCTKRM	FynTR
DNAJC7	7266	NP_003306	Y327	AVKLDDTYIKAYLRR	FRK
DOCK2	1794	NP_004937	Y481	NEYRSVVYYQVKQPR	ABL2, BLK
DOK1	1796	NP_001372	Y146	EMLENSLYSPTWEGS	BMX
DOK1	1796	NP_001372	Y362	DPKEDPIYDEPEGLA	SRC
DOK1	1796	NP_001372	Y377	PVPPQGLYDLPREPK	FRK, LYN
DOK2	9046	NP_003965	Y299	PRGQEGEYAVPFDAV	FGR
DOK6	220164	NP_689934	Y321	PLSRSSSYGFSYSSS	PTK2
DPYSL2	1808	NP_001377	Y3	-----MSYQGKKNIP	SYK
DSCAML1	57453	NP_065744	Y593	CRVIGYPYYSIKWYK	FRK
DSP	1832	NP_004406	Y1169	KAIKEKEYEIERLRV	FES
DSP	1832	NP_004406	Y2858	DATGNSSYSYSYSFS	PTK2
DTX2	113878	NP_065943	Y156	GYNYTVNYTTHTQTN	SRMS
DVL2	1856	NP_004413	Y362	WDPSPQAYFTLPRNE	SYK
DYNC1H1	1778	NP_001367	Y3379	MSNPSYNYEIVNRAS	FES
DYNC1LI2	1783	NP_006132	Y284	EKNLDLLYKYIVHKT	HCK
DYRK1A	1859	NP_001387	Y111	YKHINEVYYAKKKRR	BLK, BMX, HCK
DYRK1B	9149	NP_004705	Y63	YKHINEVYYAKKKRR	BLK, BMX, HCK
DYRK4	8798	NP_003836	Y264	EHQKVYTYIQSRFYR	ABL2
ECEL1	9427	NP_004817	Y239	LYKAQGVYSAAALFS	FRK
EEF2	1938	NP_001952	Y443	PGKKEDLYLKPIQRT	ABL2
EFNB1	1947	NP_004420	Y343	PQSPANIYYKV----	FER
EFNB2	1948	NP_004084	Y330	PQSPANIYYKV----	FER
EFS	10278	NP_005855	Y253	GGTDEGIYDVPLLGP	ABL2

EGFR	1956	NP_005219	Y915	MTFGSKPYDGIPASE	FGR
EIF2AK3	9451	NP_004827	Y481	YPYDNGYYLPYYKRE	BMX
EIF2S1	1965	NP_004085	Y150	KRPGYGAYDAFKHAV	BLK, FGR, FYN
EIF3A	8661	NP_003741	Y32	QPALDVLYDVMKSKK	HCK, SRC, YES1
EIF3C	8663	NP_003743	Y913	QQQSQTAY-------	SRMS
EIF3CL	728689	NP_001093131	Y914	QQQSQTAY-------	SRMS
EIF3I	8668	NP_003748	Y241	SAALSPNYDHVVLGG	FES
EIF3L	51386	NP_057175	Y17	SEAAYDPYAYPSDYD	FRK
EIF4A1	1973	NP_001407	Y197	RGFKDQIYDIFQKLN	BLK, FYN, FynTR
EIF4A3	9775	NP_055555	Y202	KGFKEQIYDVYRYLP	ABL2, BMX, FER, FYN, FynTR, SRC, SRMS
EIF4EBP1	1978	NP_004086	Y34	VQLPPGDYSTTPGGT	FGR
EIF4EBP2	1979	NP_004087	Y34	AAQLPHDYCTTPGGT	TNK1
EIF4H	7458	NP_071496	Y12	DTYDDRAYSSFGGGR	FGR
EIF5	1983	NP_001960	Y405	DENIEVVYSKAASVP	FYN, HCK, LYN
ELFN2	114794	NP_443138	Y743	RSKRDSTYSQLSPRH	BLK
ELMO1	9844	NP_055615	Y724	PSNYDFVYDCN----	SRMS
ELP3	55140	NP_060561	Y251	EIGVQSVYEDVARDT	BLK, FER, FRK
ELP3	55140	NP_060561	Y329	VIRGTGLYELWKSGR	BLK, FER, FGR, FYN, HCK, YES1
ENO1	2023	NP_001419	Y44	SGASTGIYEALELRD	FRK
ENO1	2023	NP_001419	Y131	VEKGVPLYRHIADLA	LYN
ENO2	2026	NP_001966	Y25	PTVEVDLYTAKGLFR	FRK
ENO2	2026	NP_001966	Y44	SGASTGIYEALELRD	FRK
ENO3	2027	NP_001967	Y44	SGASTGIYEALELRD	FRK
ENO3	2027	NP_001967	Y131	AEKGVPLYRHIADLA	ABL2, LYN
EPB41L1	2036	NP_036288	Y68	LDMEEKDYSEADGLS	SYK
EPC1	80314	NP_079485	Y175	DELIREVYEYWIKKR	FES, YES1
EPHA3	2042	NP_005224	Y779	EDDPEAAYTTRGGKI	FGR, FYN
EPHA4	2043	NP_004429	Y779	EDDPEAAYTTRGGKI	FGR, FYN
EPHA5	2044	NP_004430	Y833	EDDPEAAYTTRGGKI	FGR, FYN
EPHA6	285220	NP_001073917	Y925	EDDPEAAYTTTGGKI	FGR
EPHA7	2045	NP_004431	Y597	QEGDEELYFHFKFPG	FGR, HCK
EPHA7	2045	NP_004431	Y791	EDDPEAVYTTTGGKI	FGR
EPHB2	2048	NP_004433	Y781	DDTSDPTYTSALGGK	FYN
EPHB4	2050	NP_004435	Y987	TGGPAPQY-------	PTK2
EPN3	55040	NP_060427	Y186	SRGSPSSYNSSSSSP	PTK2
EPS8L3	79574	NP_620641	Y10	RPSSRAIYLHRKEYS	BMX
ERBB2	2064	NP_004439	Y923	MTFGAKPYDGIPARE	FGR
ERBB2	2064	NP_004439	Y1139	TCSPQPEYVNQPDVR	PTK2, YES1
ERBB2IP	55914	NP_001006600	Y884	EIGGLKIYDILSDNG	BMX
ERBB3	2065	NP_001973	Y1197	EEDEDEEYEYMNRRR	FRK
ERBB4	2066	NP_005226	Y875	LEGDEKEYNADGGKM	SYK
ERG	2078	NP_001129626	Y259	TTRPDLPYEPPRRSA	FGR
ERRFI1	54206	NP_061821	Y394	KKVSSTHYYLLPERP	FER
ESR2	2100	NP_001428	Y56	YSPAVMNYSIPSNVT	FES
ESYT2	57488	NP_065779	Y796	SEDGSDPYVRMYLLP	FRK
EVPL	2125	NP_001979	Y752	VQDAALTYQQFKNCK	ABL2
EVPL	2125	NP_001979	Y1352	RAAEDAVYELQSKRL	BMX, FER, FRK
EWSR1	2130	NP_053733	Y422	KGDATVSYEDPPTAK	FER
EWSR1	2130	NP_053733	Y661	QERRDRPY-------	SRC

F2R	2149	NP_001983	Y383	SECQRYVYSILCCKE	FES
FAF2	23197	NP_055428	Y81	ADHRIYSYVVSRPQP	BMX
FAIM3	9214	NP_005440	Y315	PRSQNNIYSACPRRA	FER, HCK, LYN, PTK2, YES1
FAM103A1	83640	NP_113640	Y104	EPYYPQQYGHYGYNQ	HCK
FAM108A1	81926	NP_112490	Y35	FLPPEATYSLVPEPE	FER
FAM120A	23196	NP_055427	Y393	GAPGQGPYPYSLSEP	TNK1
FAM125A	93343	NP_612410	Y204	LRRNDSIYEASSLYG	BMX
FAM135A	57579	NP_001099001	Y909	GYYEETDYSALDGTI	FYN
FAM175B	23172	NP_115558	Y342	QAVGSSNYASTSAGL	TNK1
FAM194B	220081	NP_872348	Y151	DYIEEVDYLGKKAYL	SYK
FAM195B	348262	NP_997251	Y41	EENVRFIYEAWQGVE	FER
FAM208A	23272	NP_001106207	Y805	GLSTDDAYEELRQKH	BLK, FYN
FAM221A	340277	NP_954587	Y38	KLFTPEEYEEYKRKV	BLK
FAM83A	84985	NP_116288	Y398	DGPPAAVYSNLGAYR	FER
FAS	355	NP_000034	Y291	LHGKKEAYDTLIKDL	SRMS
FAT1	2195	NP_005236	Y3253	GTEVLQVYAASRDIE	BMX
FAT1	2195	NP_005236	Y4356	EEKPSQPYSARESLS	FGR, SRC
FAT4	79633	NP_078858	Y4980	KDGEAEQYV------	BMX
FBP1	2203	NP_000498	Y259	DVHRTLVYGGIFLYP	FYN, SRC
FBP2	8789	NP_003828	Y259	DVHRTLVYGGIFLYP	FYN, SRC
FBXO9	26268	NP_036479	Y111	EEQNGALYEAIKFYR	FGR
FCER1G	2207	NP_004097	Y65	YEKSDGVYTGLSTRN	FER, FRK, LYN, SRC, YES1
FCER1G	2207	NP_004097	Y76	STRNQETYETLKHEK	FynTR
FCGR2A	2212	NP_001129691	Y240	AAVVALIYCRKKRIS	HCK
FCRL6	343413	NP_001004310	Y371	GKDEGVVYSVVHRTS	FER, HCK
FERMT1	55612	NP_060141	Y191	SKTMTPIYDPINGTP	FER
FERMT2	10979	NP_001128471	Y185	IYSSPGLYSKTMTPT	PTK2
FES	2242	NP_001996	Y811	RPSFSTIYQELQSIR	ABL2, BLK, HCK
FGD6	55785	NP_060821	Y696	KAYSTENYSLESQKK	FES
FGF10	2255	NP_004456	Y161	ERIEENGYNTYASFN	SRMS
FGF10	2255	NP_004456	Y164	EENGYNTYASFNWQH	YES1
FGFR1	2260	NP_075598	Y605	KDLVSCAYQVARGME	FER
FGFR1OP	11116	NP_008976	Y267	IPKPEKTYGLRKEPR	SRC
FGFR3	2261	NP_000133	Y599	KDLVSCAYQVARGME	FER
FGR	2268	NP_005239	Y34	SYGAADHYGPDPTKA	PTK2
FIP1L1	81608	NP_112179	Y113	GAPQYGSYGTAPVNL	TNK1
FIP1L1	81608	NP_112179	Y426	RSARAFPYGNVAFPH	SRC
FLI1	2313	NP_002008	Y222	SVKEDPSYDSVRRGA	FRK
FLI1	2313	NP_002008	Y263	QRPQPDPYQILGPTS	HCK
FLNA	2316	NP_001104026	Y731	KDNGNGTYSCSYVPR	ABL2, FER, FynTR
FLNA	2316	NP_001104026	Y1047	PYEVEVTYDGVPVPG	BLK, FYN, HCK
FLNB	2317	NP_001448	Y1576	HDNKDGTYAVTYIPD	SRC
FLNB	2317	NP_001448	Y1684	TAAKPGTYVIYVRFG	SRC
FLNB	2317	NP_001448	Y2502	RSSTETCYSAIPKAS	BLK, LYN
FLNC	2318	NP_001449	Y1303	SGAKTDTYVTDNGDG	PTK2
FLOT1	10211	NP_005794	Y223	YELKKAAYDIEVNTR	SRMS
FLOT1	10211	NP_005794	Y238	RAQADLAYQLQVAKT	SYK
FLT1	2321	NP_002010	Y1333	DYNSVVLYSTPPI--	BLK
FLT3	2322	NP_004110	Y166	SIRNTLLYTLRRPYF	BMX, FER
FLT3	2322	NP_004110	Y401	QKGLDNGYSISKFCN	ABL2

FN1	2335	NP_997647	Y2444	TGQSYNQYSQRYHQR	ABL2
FN3K	64122	NP_071441	Y263	PRSFFTAYHRKIPKA	PTK2
FNBP1	23048	NP_055848	Y500	ARRQSGLYDSQNPPT	PTK2
FNDC3B	64778	NP_073600	Y265	NDSDLQEYELEVKRV	BMX
FNTB	2342	NP_002019	Y300	NKLVDGCYSFWQAGL	BLK, FGR
FOXK1	221937	NP_001032242	Y219	ASPLRPLYPQISPLK	FES
FOXL2	668	NP_075555	Y258	PAASYGPYTRVQSMA	FER, FGR, TNK1
FRG1	2483	NP_004468	Y4	----MAEYSYVKSTK	SYK
FRMD4A	55691	NP_060497	Y620	ESSLDEPYEKVKKRS	FRK, HCK, SRC, YES1
FRS2	10818	NP_006645	Y306	PSVNKLVYENINGLS	FER, FES
FRS2	10818	NP_006645	Y349	RRTALLNYENLPSLP	FES, SYK
FUBP1	8880	NP_003893	Y58	TSLNSNDYGYGGQKR	FGR
FUBP1	8880	NP_003893	Y625	YYRQQAAYYAQTSPQ	PTK2
FUBP1	8880	NP_003893	Y626	YRQQAAYYAQTSPQG	PTK2, SRMS, TNK1
FYB	2533	NP_001456	Y559	GRTARGSYGYIKTTA	PTK2
FYB	2533	NP_001456	Y571	TTAVEIDYDSLKLKK	FYN, SRC
FYN	2534	NP_002028	Y420	RLIEDNEYTARQGAK	FGR, FRK
FZR1	51343	NP_001129670	Y91	NGKDGLAYSALLKNE	BMX
GAB1	2549	NP_997006	Y307	MRHVSISYDIPPTPG	FES
GAB2	9846	NP_036428	Y266	TEFRDSTYDLPRSLA	LYN
GAB2	9846	NP_036428	Y409	RASSCETYEYPQRGG	FES
GAGE1	2543	NP_001459	Y9	SWRGRSTYYWPRPRR	PTK2
GAGE4	2576	NP_001465	Y9	SWRGRSTYYWPRPRR	PTK2
GAGE5	2577	NP_001466	Y9	SWRGRSTYYWPRPRR	PTK2
GAGE6	2578	NP_001467	Y9	SWRGRSTYYWPRPRR	PTK2
GAGE7	2579	NP_066946	Y9	SWRGRSTYYWPRPRR	PTK2
GAR1	54433	NP_061856	Y97	TDENKVPYFNAPVYL	SYK
GAS2L2	246176	NP_644814	Y801	RIPRPLAYVFLGPAR	TNK1
GATA6	2627	NP_005248	Y282	AAREPGGYAAAGSGG	FRK, TNK1
GBAS	2631	NP_001474	Y187	PRSGPNIYELRSYQL	BMX, FER, FRK
GCET2	257144	NP_689998	Y80	QDNVDQTYSEELCYT	FYN
GCET2	257144	NP_689998	Y86	TYSEELCYTLINHRV	BMX, FGR
GCET2	257144	NP_689998	Y128	LGGTETEYSLLHMPS	BMX
GDI2	2665	NP_001485	Y203	DYLDQPCYETINRIK	YES1
GEMIN5	25929	NP_056280	Y1053	YLGATCAYDAAKVLA	FER, SYK
GGA2	23062	NP_055859	Y269	QEALQVVYERCEKLR	YES1
GIT1	28964	NP_001078923	Y224	ERLVECQYELTDRLA	SRC
GIT1	28964	NP_001078923	Y554	ELEDDAIYSVHVPAG	FYN
GLB1	2720	NP_000395	Y294	EAVASSLYDILARGA	FER, FynTR, HCK
GLDN	342035	NP_861454	Y382	VVYNNSLYYHKGGSN	SYK
GLUL	2752	NP_002056	Y185	AHYRACLYAGVKIAG	FYN
GMPR2	51292	NP_057660	Y336	GIRSTCTYVGAAKLK	FYN
GNAI3	2773	NP_006487	Y354	NLKECGLY-------	SRC
GOLGA4	2803	NP_002069	Y1551	LNEVLKNYNQQKDIE	SYK
GOLGA4	2803	NP_002069	Y2148	KKYEKNVYATTVGTP	FER
GPCPD1	56261	NP_062539	Y608	LGVNGLIYDRIYDWM	FES
GPD2	2820	NP_000399	Y600	ETARKFLYYEMGYKS	SRC
GPKOW	27238	NP_056513	Y159	TVPEEANYEAVPVEA	SYK
GPR155	151556	NP_689742	Y451	ILVFVLLYSSLYSTY	HCK
GPRC5A	9052	NP_003970	Y320	GDTLYAPYSTHFQLQ	FynTR
GPRC5C	55890	NP_071319	Y403	AKRPVSPYSGYNGQL	PTK2
GPX1	2876	NP_000572	Y98	EEILNSLKYVRPGGG	BLK

GRASP	160622	NP_859062	Y237	VVKDPSIYDTLESVR	BLK, FynTR, HCK
GRHPR	9380	NP_036335	Y255	VVNQDDLYQALASGK	FynTR, HCK, LYN
GRIA3	2892	NP_000819	Y460	GYCVDLAYEIAKHVR	ABL2, BLK, FGR, FYN
GRIA3	2892	NP_000819	Y499	GMVGELVYGRADIAV	FYN
GRIA3	2892	NP_000819	Y877	PATNTQNYATYREGY	FER, LYN
GRIA3	2892	NP_000819	Y887	YREGYNVYGTESVKI	FYN
GRM1	2911	NP_000829	Y1111	KLLQEYVYEHEREGN	FynTR
GSDMD	79792	NP_079012	Y158	RSRGDNVYVVTEVLQ	HCK, SRC
GSR	2936	NP_000628	Y451	STSFTPMYHAVTKRK	YES1
GSTO1	9446	NP_004823	Y239	NSPEACDYGL-----	SRC
GSTP1	2950	NP_000843	Y4	----MPPYTVVYFPV	SRC
GTF3C5	9328	NP_001116295	Y347	AKRSTYNYSLPITVK	FES
HAVCR2	84868	NP_116171	Y265	IRSEENIYTIEENVY	SRC
HBB	3043	NP_000509	Y131	TPPVQAAYQKVVAGV	FYN
HBS1L	10767	NP_006611	Y58	EPVEEYDYEDLKESS	FER
HCFC2	29915	NP_037452	Y560	KSEVDETYALPATKI	BLK, FRK
HCK	3055	NP_002101	Y522	YTATESQYQQQP---	ABL2, FRK
HCN2	610	NP_001185	Y795	RAPRTSPYGGLPAAP	TNK1
HDLBP	3069	NP_005327	Y582	ADLVENSYSISVPIF	BMX
HEATR4	399671	NP_976054	Y729	NLMQRDPYWKIKAFA	YES1
HEBP1	50865	NP_057071	Y144	GYAKEADYVAQATRL	PTK2
HEBP2	23593	NP_055135	Y179	KVYYTAGYNSPVKLL	SRMS
HEG1	57493	NP_065784	Y1350	RNGLYPAYTGLPGSR	FGR, SYK
HELB	92797	NP_387467	Y112	QVQGFPSYFLQSDMS	PTK2
HELZ	9931	NP_055692	Y1353	LPAPHAQYAIPNRHF	TNK1
HEPACAM2	253012	NP_001034461	Y412	ALDDFGIYEFVAFPD	YES1
HIST1H2BB	3018	NP_066406	Y41	RKESYSIYVYKVLKQ	BMX
HIST1H4I	8294	NP_003486	Y52	KRISGLIYEETRGVL	BMX
HIST1H4I	8294	NP_003486	Y89	VTAMDVVYALKRQGR	BLK, FynTR, LYN
HIVEP2	3097	NP_006725	Y638	GDRVGYDYDVCRKPY	FER, FYN
HIVEP3	59269	NP_078779	Y1743	GYKSNEEYVYVRGRG	BMX, FES, FRK
HIVEP3	59269	NP_078779	Y1745	KSNEEYVYVRGRGRG	FES, SRC
HNRNPA0	10949	NP_006796	Y180	AVPKEDIYSGGGGGG	FYN
HNRNPA1	3178	NP_112420	Y341	GGRSSGPYGGGGQYF	TNK1
HNRNPA1	3178	NP_112420	Y347	PYGGGGQYFAKPRNQ	TNK1
HNRNPA1	3178	NP_112420	Y357	KPRNQGGYGGSSSSS	TNK1
HNRNPA2B1	3181	NP_112533	Y262	NFGGSPGYGGGRGGY	TNK1
HNRNPA2B1	3181	NP_112533	Y336	GPYGGGNYGPGGSGG	TNK1
HNRNPA2B1	3181	NP_112533	Y347	GSGGSGGYGGRSRY-	PTK2
HNRNPA3	220988	NP_919223	Y360	GRSSGSPYGGGYGSG	FGR
HNRNPA3	220988	NP_919223	Y373	SGGGSGGYGSRRF--	PTK2
HNRNPAB	3182	NP_112556	Y332	HQNNYKPY-------	SRMS
HNRNPF	3185	NP_004957	Y240	PGAYSTGYGGYEEYS	PTK2
HNRNPF	3185	NP_004957	Y243	YSTGYGGYEEYSGLS	SRC
HNRNPF	3185	NP_004957	Y246	GYGGYEEYSGLSDGY	BLK
HNRNPF	3185	NP_004957	Y306	KATENDIYNFFSPLN	YES1
HNRNPH2	3188	NP_062543	Y243	YGGGYGGYDDYGGYN	FER
HNRNPH3	3189	NP_036339	Y331	GGGGSGGYYGQGGMS	PTK2
HNRNPUL1	11100	NP_008971	Y717	QPPPPPSYSPARNPP	TNK1
HNRNPUL2	221092	NP_001073027	Y660	NRSRGQGYVGGQRRG	TNK1
HNRNPUL2	221092	NP_001073027	Y743	YYRNYYGYQGYR---	LYN, PTK2, SRMS

HOXD8	3234	NP_062458	Y10	SYFVNPLYSKYKAAA	ABL2, FER, LYN
HPSE2	60495	NP_068600	Y302	IYSRASLYGPNIGRP	PTK2
HRAS	3265	NP_005334	Y157	QGVEDAFYTLVREIR	FRK
HRH1	3269	NP_000852	Y301	DREVDKLYCFPLDIV	LYN
HS2ST1	9653	NP_036394	Y78	EEDMVIIYNRVPKTA	YES1
HSPA1A	3303	NP_005336	Y611	NPIISGLYQGAGGPG	TNK1
HSPA4	3308	NP_002145	Y89	AEKSNLAYDIVQLPT	BMX
HSPA4	3308	NP_002145	Y336	KLKKEDIYAVEIVGG	FYN
HSPA8	3312	NP_006588	Y15	GIDLGTTYSCVGVFQ	FYN, YES1
HSPH1	10808	NP_006635	Y89	KEKENLSYDLVPLKN	FER, FES
HTATSF1	27336	NP_055315	Y634	KEEEEDTYEKVFDDE	HCK
HYLS1	219844	NP_659451	Y48	REAQSIQYDPYSKAS	LYN
ICK	22858	NP_057597	Y156	EIRSKPPYTDYVSTR	FER
IFT74	80173	NP_079379	Y572	TKSQESDYQPIKKNV	BMX
IGF1R	3480	NP_000866	Y1166	DIYETDYYRKGGKGL	LYN
IKZF1	10320	NP_006051	Y293	KGLSDTPYDSSASYE	SRMS
IL22RA1	58985	NP_067081	Y301	SLAQPVQYSQIRVSG	ABL2
IL6ST	3572	NP_002175	Y759	NTSSTVQYSTVVHSG	BMX, LYN
IL6ST	3572	NP_002175	Y814	GILPRQQYFKQNCSQ	PTK2
ILF3	3609	NP_060090	Y365	QIPPSTTYAITPMKR	FGR
ILF3	3609	NP_060090	Y583	NKKVAKAYAALAALE	SYK
ILF3	3609	NP_060090	Y840	NPGSHGGYGGGSGGG	TNK1
INPP4A	3631	NP_001127696	Y972	YRPPEGTYGKVET--	SRC
INPP5D	3635	NP_001017915	Y1022	EMFENPLYGSLSSFP	FYN
INPPL1	3636	NP_001558	Y671	ERGSRDTYAWHKQKP	FYN
INPPL1	3636	NP_001558	Y831	TVKSMDGYESYGECV	LYN
INPPL1	3636	NP_001558	Y1135	KTLSEVDYAPAGPAR	FER
INSR	3643	NP_000199	Y1190	DIYETDYYRKGGKGL	LYN
INTS4	92105	NP_291025	Y686	WNVAAPLYLKQSDLA	SYK
IQGAP1	8826	NP_003861	Y654	RSPDVGLYGVIPECG	BLK, SRC
IQGAP1	8826	NP_003861	Y855	ANKARDDYKTLINAE	SYK
IQSEC2	23096	NP_001104595	Y933	RDLLVGIYQRIQGRE	FES
IREB2	3658	NP_004127	Y40	TKYDVLPYSIRVLLE	BMX, SRC
IRS2	8660	NP_003740	Y823	CGGDSDQYVLMSSPV	FRK
ISY1	57461	NP_065752	Y99	KELGGPDYGKVGPKM	FYN
ITGA3	3675	NP_002195	Y1051	TERLTDDY-------	SYK
ITGB1BP1	9270	NP_004754	Y127	KVSTSDQYDVLHRHA	HCK
ITGB7	3695	NP_000880	Y753	YRLSVEIYDRREYSR	BLK
ITPR1	3708	NP_001093422	Y482	KLLEDLVYFVTGGTN	FER, FGR, SRC
ITPR1	3708	NP_001093422	Y2616	KVKDSTEYTGPESYV	FGR
ITPR2	3709	NP_002214	Y2607	KVKDPTEYTGPESYV	FGR
ITSN2	50618	NP_006268	Y39	NLKPSGGYITGDQAR	TNK1
ITSN2	50618	NP_006268	Y968	REEPEALYAAVNKKP	FRK
JAK1	3716	NP_002218	Y1034	AIETDKEYYTVKDDR	FRK
JUP	3728	NP_002221	Y644	RNEGTATYAAAVLFR	FYN
KAT2B	8850	NP_003875	Y729	PRDPDQLYSTLKSIL	FRK, FYN, LYN
KCNAB2	8514	NP_003627	Y25	TGSPGMIYSTRYGSP	PTK2
KCNC3	3748	NP_004968	Y574	PQPGSPNYCKPDPPP	FER, TNK1
KCND2	3751	NP_036413	Y134	EIIGDCCYEEYKDRR	BMX
KCNQ2	3785	NP_742105	Y372	RTVTVPMYSSQTQTY	YES1
KDELC2	143888	NP_714916	Y380	VDGTVAAYRYPYLML	SRMS
KDM1A	23028	NP_001009999	Y136	NLSEDEYYSEEERNA	HCK

KHDRBS1	10657	NP_006550	Y380	DTYAEQSYEGYEGYY	SRC
KHDRBS1	10657	NP_006550	Y443	REHPYGRY-------	FES
KIAA0368	23392	NP_001073867	Y274	VTNFTIIYVKMGYPR	HCK, YES1
KIAA1217	56243	NP_062536	Y244	KDESRNVYYELNDVR	BMX
KIAA1217	56243	NP_062536	Y435	AIRSASAYCNPSMQA	PTK2
KIAA1217	56243	NP_062536	Y520	KEPGTLVYIEKPRSA	LYN
KIAA1430	57587	NP_065878	Y292	ENVSQEIYEDVEDLK	FRK
KIAA1462	57608	NP_065899	Y340	PGLEPPVYVPPPSYR	HCK
KIAA1522	57648	NP_065939	Y584	TLSPSSGYSSQSGTP	PTK2
KIAA1522	57648	NP_065939	Y737	PPSPPPSYHPPPPPT	PTK2, TNK1
KIAA1586	57691	NP_065982	Y448	KIFIDKIYSIYHQPN	FYN
KIFAP3	22920	NP_055785	Y787	FRPDEPYYYGYGS--	SRC, SYK
KIR2DL2	3803	NP_055034	Y126	DIVITGLYEKPSLSA	YES1
KIT	3815	NP_000213	Y570	INGNNYVYIDPTQLP	FRK
KLC1	3831	NP_001123579	Y360	GKYEEVEYYYQRALE	BMX
KLC2	64837	NP_073733	Y345	GKAEEVEYYYRRALE	BMX
KLC4	89953	NP_958931	Y223	TAAQQGGYEIPARLR	ABL2
KLC4	89953	NP_958931	Y611	KRAASLNYLNQPSAA	TNK1
KLF3	51274	NP_057615	Y38	PNKYGVIYSTPLPEK	SRMS
KLHDC9	126823	NP_689579	Y164	QRRYGSIYTLRLDPS	BMX
KMO	8564	NP_003670	Y206	IPPKNGDYAMEPNYL	SYK
KRI1	65095	NP_075384	Y518	DEYYRLDYEDIIDDL	FES, SYK
KRT1	3848	NP_006112	Y639	VKFVSTTYSGVTR--	FER
KRT3	3850	NP_476429	Y393	KAEAEALYQTKLGEL	SYK
KRT39	390792	NP_998821	Y379	LERQNQEYEILLDVK	BMX
KRT4	3851	NP_002263	Y389	IAEVRAQYEEIAQRS	FRK
KRT4	3851	NP_002263	Y404	KAEAEALYQTKVQQL	BMX
KRT5	3852	NP_000415	Y66	GYGSRSLYNLGGSKR	LYN
KRT5	3852	NP_000415	Y361	RTEAESWYQTKYEEL	SYK
KRT6A	3853	NP_005545	Y83	SCAISGGYGSRAGGS	PTK2
KRT6A	3853	NP_005545	Y356	RAEAESWYQTKYEEL	SYK
KRT6B	3854	NP_005546	Y83	SCAISGGYGSRAGGS	PTK2
KRT6B	3854	NP_005546	Y356	RAEAESWYQTKYEEL	SYK
KRT6C	286887	NP_775109	Y83	SCAISGGYGSRAGGS	PTK2
KRT6C	286887	NP_775109	Y356	RAEAESWYQTKYEEL	SYK
KRT8	3856	NP_002264	Y228	INFLRQLYEEEIREL	SYK
KRT9	3857	NP_000217	Y485	RGGSGGSYGRGSRGG	TNK1
LAMA4	3910	NP_001098676	Y1324	SSVSPTRYELIVDKS	FES
LAMTOR1	55004	NP_060377	Y140	SRIAAYAYSALSQIR	ABL2
LANCL2	55915	NP_061167	Y198	DLPDELLYGRAGYLY	HCK
LAPTM4B	55353	NP_060877	Y314	AKEPPPPYVSA----	FRK
LAPTM5	7805	NP_006753	Y259	GGPAPPPYSEV----	HCK, TNK1
LARP1	23367	NP_056130	Y700	TVPESPNYRNTRTPR	LYN, PTK2
LARP1	23367	NP_056130	Y785	GTPTVGSYGCTPQSL	PTK2, TNK1
LARP4B	23185	NP_055970	Y596	SCAVSATYERSPSPA	PTK2
LASP1	3927	NP_006139	Y171	IPTSAPVYQQPQQQP	ABL2
LAT2	7462	NP_115853	Y118	IDPIAMEYYNWGRFS	PTK2
LATS2	26524	NP_055387	Y286	TPPETGGYASLPTKG	LYN
LCP1	3936	NP_002289	Y430	ALVIFQLYEKIKVPV	FRK
LEMD2	221496	NP_851853	Y104	AYATPGAYGDIRPSA	PTK2
LGALS3	3958	NP_002297	Y107	AYPATGPYGAPAGPL	PTK2, TNK1
LGALS3	3958	NP_002297	Y118	AGPLIVPYNLPLPGG	SRMS

LILRB1	10859	NP_001075106	Y534	DAQEENLYAAVKHTQ	ABL2, SRC
LILRB1	10859	NP_001075106	Y646	SPAVPSIYATLAIH-	HCK
LILRB2	10288	NP_005865	Y533	DAQEENLYAAVKDTQ	ABL2, SRC
LILRB2	10288	NP_005865	Y592	PPAEPSIYATLAIH-	HCK
LILRB4	11006	NP_006838	Y360	EDPQAVTYAKVKHSR	ABL2
LIME1	54923	NP_060276	Y145	CAGLEATYSNVGLAA	FER
LIME1	54923	NP_060276	Y167	ASPVVAEYARVQKRK	FRK
LIME1	54923	NP_060276	Y235	ALAGDLAYQTLPLRA	ABL2, FGR, FYN
LIME1	54923	NP_060276	Y254	SGPLENVYESIRELG	FES, FRK, LYN
LITAF	9516	NP_004853	Y32	ETVAVNSYYPTPPAP	FER, TNK1
LMO7	4008	NP_056667	Y38	NKSRQPSYVPAPLRK	BMX
LPHN1	22859	NP_001008701	Y1441	RNPLQGYYQVRRPSH	FYN
LPP	4026	NP_005569	Y244	APSSGQIYGSGPQGY	FYN, LYN
LPP	4026	NP_005569	Y251	YGSGPQGYNTQPVPV	TNK1
LPP	4026	NP_005569	Y301	GRYYEGYYAAGPGYG	FGR, FYN
LPXN	9404	NP_001137467	Y77	NIQELNVYSEAQEPK	BMX
LRP1B	53353	NP_061027	Y3253	ADRLSLIYSWHAITD	FYN, FynTR
LRRC25	126364	NP_660299	Y284	DLASQPVYCNLQSLG	LYN, YES1
LRRC58	116064	NP_001093148	Y280	RNISYTPYDLPGNLL	SRMS
LRRFIP2	9209	NP_006300	Y300	DEKSDKQYAENYTRP	LYN
LRRFIP2	9209	NP_006300	Y304	DKQYAENYTRPSSRN	ABL2
LSR	51599	NP_991403	Y616	EEEEEAYYPPAPPPY	BMX
LSR	51599	NP_991403	Y623	YPPAPPPYSETDSQA	TNK1
LTBP4	8425	NP_001036009	Y1449	YPPPALPYDPYPPPP	SRC
LTK	4058	NP_002335	Y676	RDIYRASYYRRGDRA	PTK2
LY6G6F	259215	NP_001003693	Y281	FKPEIQVYENIHLAR	FER
LY9	4063	NP_002339	Y626	EESSATIYCSIRKPQ	FynTR, HCK, SRC
LYN	4067	NP_002341	Y316	VTREEPIYIITEYMA	SRC
LYN	4067	NP_002341	Y508	YTATEGQYQQQP---	ABL2, FRK
LYRM4	57128	NP_065141	Y28	KRFSAYNYRTYAVRR	SRMS
LZTS1	11178	NP_066300	Y295	ELASSLAYEERPRRC	FGR
MAG	4099	NP_002352	Y620	LTEELAEYAEIRVK-	SYK
MAGEC2	51438	NP_057333	Y259	VLNAVGVYAGREHFV	BLK, SRC
MAGED1	9500	NP_001005333	Y182	PKGPNAAYDFSQAAT	FER
MAGI2	9863	NP_036433	Y362	EKIDDPIYGTYYVDH	FYN
MANBA	4126	NP_005899	Y481	RPIYIKDYVTLYVKN	FES
MAP1B	4131	NP_005900	Y1174	SSLYSQEYSKPADVT	FES
MAP1B	4131	NP_005900	Y1543	EGVAEDTYSHMEGVA	FRK
MAP1B	4131	NP_005900	Y1796	PRESSPLYSPTFSDS	HCK
MAP1B	4131	NP_005900	Y1906	SDVGGYYYEKIERTT	FER, FynTR, HCK
MAP1B	4131	NP_005900	Y1955	RTPEEGGYSYDISEK	SYK
MAP3K3	4215	NP_976226	Y186	AGDINTIYQPPEPRS	PTK2
MAP4	4134	NP_002366	Y110	FLEEKMAYQEYPNSQ	SYK
MAP4K3	8491	NP_003609	Y366	LDSSEEIYYTARSNL	BMX
MAP7D3	79649	NP_078873	Y256	VTNYVMQYVTVPLRK	SRC
MAP9	79884	NP_001034669	Y12	VFSTTLAYTKSPKVT	FER
MAPKAPK3	7867	NP_004626	Y208	TPCYTPYYVAPEVLG	ABL2, HCK
MARK2	2011	NP_001034558	Y613	RDQQNLPYGVTPASP	FGR, PTK2, TNK1
MATR3	9782	NP_061322	Y250	YHKFDSEYERMGRGP	FRK
MATR3	9782	NP_061322	Y526	PYGKIKNYILMRMKS	SRMS
MB21D1	115004	NP_612450	Y248	EYSNTRAYYFVKFKR	FGR
MBD1	4152	NP_056671	Y403	GAGSPPPYRRRKRPS	TNK1

MDH1	4190	NP_005908	Y210	QGKEVGVYEALKDDS	BMX, FRK
MDM2	4193	NP_002383	Y411	STSSSIIYSSQEDVK	YES1
MDN1	23195	NP_055426	Y2022	PYDVQLGYSVLSRGS	FER
MDN1	23195	NP_055426	Y5163	DAYDAQTYDVASKEQ	FYN
MED9	55090	NP_060489	Y63	RAREEENYSFLPLVH	FER, FES
MELK	9833	NP_055606	Y427	LKNKENVYTPKSAVK	BMX
MEPCE	56257	NP_062552	Y418	RKFQYGNYCKYYGYR	TNK1
MGRN1	23295	NP_056061	Y416	PLYEEITYSGISDGL	HCK
MINA	84864	NP_694822	Y76	DDPALATYYGSLFKL	SYK
MINA	84864	NP_694822	Y77	DPALATYYGSLFKLT	SRC
MLF1IP	79682	NP_078905	Y68	NEKDEETYETFDPPL	FYN
MLLT4	4301	NP_001035089	Y1230	TQTYTREYFTFPASK	FGR
MMP16	4325	NP_005932	Y52	VWLQKYGYLPPTDPR	SRMS
MMP3	4314	NP_002413	Y42	QKYLENYYDLKKDVK	BMX, FRK
MPDZ	8777	NP_003820	Y1143	SELQNTAYSNWNQPR	FER, PTK2
MPLKIP	136647	NP_619646	Y13	FRPPTPPYPGPGGGG	TNK1
MPP1	4354	NP_002427	Y316	KFVYPVPYTTRPPRK	SRMS
MPP2	4355	NP_005365	Y115	DSVASKTYETPPPSP	FES
MPP2	4355	NP_005365	Y339	DRHELLIYEEVARMP	SRC, SYK
MPP6	51678	NP_057531	Y327	DRHEIQIYEEVAKMP	FES, SRC
MPP7	143098	NP_775767	Y354	DTADVPTYEEVTPYR	BLK, SRC, YES1
MPZL1	9019	NP_003944	Y241	SHQGPVIYAQLDHSG	ABL2
MRPL24	79590	NP_663781	Y216	KYKKVYWY-------	FES, SRC, SRMS, SYK
MRRF	92399	NP_620132	Y47	GHRQYMAYSAVPVRH	ABL2
MSN	4478	NP_002435	Y116	GILNDDIYCPPETAV	LYN
MTA2	9219	NP_004730	Y483	RRAARRPYAPINANA	TNK1
MTMR6	9107	NP_004676	Y614	PAVVSLEYGVARMTC	FGR
MUC1	4582	NP_002447	Y247	SSTDRSPYEKVSAGN	FER
MUC1	4582	NP_002447	Y261	NGGSSLSYTNPAVAA	TNK1
MUC13	56667	NP_149038	Y511	SSMPRPDY-------	FES
MVP	9961	NP_059447	Y58	VTVPPRHYCTVANPV	TNK1
MX2	4600	NP_002454	Y697	RILKERIYRLTQARH	BMX
MYBPC3	4607	NP_000247	Y373	QKKLEPAYQVSKGHK	TNK1
MYH1	4619	NP_005954	Y413	RVKVGNEYVTKGQTV	FRK
MYH13	8735	NP_003793	Y412	RVKVGNEYVTKGQNV	BMX, FRK
MYH2	4620	NP_060004	Y413	RVKVGNEYVTKGQTV	FRK
MYH3	4621	NP_002461	Y411	RVKVGNEYVTKGQTV	FRK
MYH6	4624	NP_002462	Y351	SEEKAGVYKLTGAIM	BMX
MYH7	4625	NP_000248	Y410	RVKVGNEYVTKGQNV	BMX, FRK
MYH8	4626	NP_002463	Y413	RVKVGNEYVTKGQTV	FRK
MYLK	4638	NP_444253	Y1635	VAPEVINYEPIGYAT	FER, FES
MYO10	4651	NP_036466	Y585	ESRFDFIYDLFEHVS	FynTR, SRC
MYO10	4651	NP_036466	Y1128	YRCSVGTYNSSGAYR	PTK2
MYO1E	4643	NP_004989	Y950	SGTQNANYPVRAAPP	PTK2
MYO1E	4643	NP_004989	Y989	RSNQKSLYTSMARPP	HCK
MYT1	4661	NP_004526	Y428	DTVRKSYYSKDPSRA	PTK2
N4BP3	23138	NP_055926	Y83	PRNEPADYATLYYRE	SRMS
NAA15	80155	NP_476516	Y66	LGKKEEAYELVRRGL	BLK, FER
NASP	4678	NP_002473	Y148	EELREQVYDAMGEKE	SRC
NAV2	89797	NP_892009	Y1579	ADGQYDPYTDSRFRN	ABL2
NCF1	653361	NP_000256	Y48	YRRFTEIYEFHKTLK	FynTR
NCK2	8440	NP_003572	Y50	NAANRTGYVPSNYVE	BMX

NCKIPSD	51517	NP_057537	Y161	LGADGGLYQIPLPSS	ABL2
NCL	4691	NP_005372	Y462	DGRSISLYYTGEKGQ	LYN
NCOR2	9612	NP_006303	Y1511	RALERACYEESLKSR	BMX
NDC80	10403	NP_006092	Y458	VKYRAQVYVPLKELL	BMX
NDFIP2	54602	NP_061953	Y167	TTSDTEVYGEFYPVP	FYN, HCK
NDFIP2	54602	NP_061953	Y177	FYPVPPPYSVATSLP	FER
NEB	4703	NP_004534	Y1824	DYKYKKAYEQAKGKH	SYK
NEB	4703	NP_004534	Y4918	DAVYHYDYVHSVRGK	FES
NEDD9	4739	NP_006394	Y106	AAPRDTIYQVPPSYQ	ABL2, FynTR, LYN
NEDD9	4739	NP_006394	Y112	IYQVPPSYQNQGIYQ	PTK2
NEDD9	4739	NP_006394	Y132	GTQEQEVYQVPPSVQ	ABL2
NEDD9	4739	NP_006394	Y177	SRYQKDVYDIPPSHT	HCK
NEDD9	4739	NP_006394	Y189	SHTTQGVYDIPPSSA	ABL2
NEDD9	4739	NP_006394	Y261	DLRPEGVYDIPPTCT	FER, FES
NEDD9	4739	NP_006394	Y317	VGSQNDAYDVPRGVQ	PTK2
NEFL	4747	NP_006149	Y43	ARSAYSSYSAPVSSS	PTK2, SYK
NETO1	81832	NP_620416	Y417	VADDFENYHKLRRSS	ABL2
NFAM1	150372	NP_666017	Y267	DGELNLVYENL----	BLK, FES, LYN
NFIX	4784	NP_002492	Y253	ADLESPSYYNINQVT	PTK2
NGEF	25791	NP_062824	Y179	IEQIGLLYQEYRDKS	ABL2
NGFR	4804	NP_002498	Y336	LKGDGGLYSSLPPAK	FYN
NIPSNAP1	8508	NP_003625	Y185	PRMGPNIYELRTYKL	BLK
NOLC1	9221	NP_004732	Y289	MKNKPGPYSSVPPPS	SRC
NOP58	51602	NP_057018	Y272	SEYRTQLYEYLQNRM	BMX, FES
NOP58	51602	NP_057018	Y342	TPKYGLIYHASLVGQ	ABL2
NOS2	4843	NP_000616	Y868	ALCQPSEYSKWKFTN	ABL2, YES1
NPAS4	266743	NP_849195	Y39	ADKVRLSYLHIMSLA	SYK
NPHP1	4867	NP_000263	Y46	PNKRQHIYQRCIQLK	FYN
NPHP1	4867	NP_000263	Y722	FDLSEQTYDFLGEMR	FER
NPHS1	4868	NP_004637	Y1193	SGAWGPLYDEVQMGP	BMX
NR2C1	7181	NP_003288	Y208	AASTEKIYIRKDLRS	FRK
NRAP	4892	NP_932326	Y523	YEKNKLNYTLPQDVP	LYN
NRAS	4893	NP_002515	Y157	QGVEDAFYTLVREIR	FRK
NRK	203447	NP_940867	Y984	VDDVNNNYYEAPSCP	ABL2
NRK	203447	NP_940867	Y985	DDVNNNYYEAPSCPR	FER, FynTR
NRK	203447	NP_940867	Y1191	EVNVNPLYVSPACKK	FRK, HCK
NRP1	8829	NP_003864	Y920	KLNTQSTYSEA----	BLK, FRK, HCK
NRXN1	9378	NP_001129131	Y1076	GGVAKETYKSLPKLV	HCK
NSF	4905	NP_006169	Y259	HVKGILLYGPPGCGK	FER
NSFL1C	55968	NP_057227	Y155	RPFAGGGYRLGAAPE	TNK1
NTRK1	4914	NP_002520	Y680	RDIYSTDYYRVGGRT	FES
NTRK2	4915	NP_006171	Y722	RDVYSTDYYRVGGHT	FES
NTRK2	4915	NP_006171	Y833	LAKASPVYLDILG--	FES
NTRK3	4916	NP_001012338	Y516	PVIENPQYFRQGHNC	PTK2
NTRK3	4916	NP_001012338	Y834	LGKATPIYLDILG--	FES, SYK
NTS	4922	NP_006174	Y153	YILKRQLYENKPRRP	FES
NUDCD2	134492	NP_660309	Y145	GAEISGNYTKGGPDF	TNK1
NUMA1	4926	NP_006176	Y1836	DSANSSFYSTRSAPA	PTK2
NUP214	8021	NP_005076	Y1265	GGGSKPSYEAIPESS	FER, SYK
NUP93	9688	NP_055484	Y391	DPYKRAVYCIIGRCD	FynTR, SRC
NYAP2	57624	NP_065915	Y268	DDQSEAVYEEMKYPI	SRC
OBSCN	84033	NP_001092093	Y5864	CALLEQAYAVVSALP	FER

ODF3B	440836	NP_001014440	Y23	RGPIAAHYGGPGPKY	PTK2, TNK1
ODF3L2	284451	NP_872383	Y49	NGSGPGLYVLPSTVG	FRK, HCK
ODF3L2	284451	NP_872383	Y70	TRVASPAYSLVRRPS	FES
ODF3L2	284451	NP_872383	Y107	GRSCTPAYSMQGRAK	SYK
ORC3	23595	NP_862820	Y528	GLQKTDLYHLQKSLL	ABL2
ORC6	23594	NP_055136	Y232	DEDLTQDYEEWKRKI	FYN
OTUD5	55593	NP_060072	Y181	GYNSEDEYEAAAARI	FRK
PABPC1	26986	NP_002559	Y54	ITRRSLGYAYVNFQQ	TNK1
PABPC1	26986	NP_002559	Y56	RRSLGYAYVNFQQPA	TNK1
PABPC1	26986	NP_002559	Y116	SIDNKALYDTFSAFG	YES1
PABPC1	26986	NP_002559	Y364	IVATKPLYVALAQRK	HCK
PABPC3	5042	NP_112241	Y364	IVATKPLYVALAQRK	HCK
PABPC4	8761	NP_001129125	Y54	ITRRSLGYAYVNFQQ	TNK1
PABPC4	8761	NP_001129125	Y56	RRSLGYAYVNFQQPA	TNK1
PABPC4	8761	NP_001129125	Y116	SIDNKALYDTFSAFG	YES1
PACS1	55690	NP_060496	Y370	EEDLDELYDSLEMYN	BLK, FRK
PACSIN3	29763	NP_057307	Y372	GVRVRALYDYAGQEA	FYN, SYK
PAG1	55824	NP_060910	Y163	GLGMEGPYEVLKDSS	YES1
PAG1	55824	NP_060910	Y341	LTVPESTYTSIQGDP	FRK
PAG1	55824	NP_060910	Y387	SEEPEPDYEAIQTLN	FES
PALM2	114299	NP_443749	Y214	DDGTKVVYEVRSGGT	BLK, FER
PALMD	54873	NP_060204	Y140	EESIEDIYANIPDLP	YES1
PAPSS1	9061	NP_005434	Y30	QRATNVTYQAHHVSR	FYN
PAPSS2	9060	NP_001015880	Y20	QKSTNVVYQAHHVSR	FYN
PAQR3	152559	NP_001035292	Y45	GSLKDNPYITDGYRA	FGR
PARD3	56288	NP_062565	Y489	IGGSAPIYVKNILPR	PTK2
PARD3	56288	NP_062565	Y1321	KKVQDPSYAPPKGPF	FER, TNK1
PARD3B	117583	NP_689739	Y938	RDHLEGLYAKVNKPY	BMX, SRC
PARK7	11315	NP_009193	Y67	DAKKEGPYDVVVLPG	FES, SRC, YES1
PBRM1	55193	NP_060635	Y1257	KVVDDEIYYFRKPIV	FYN, HCK
PCBP2	5094	NP_005007	Y236	AYTIQGQYAIPQPDL	ABL2
PCDH1	5097	NP_115796	Y897	KKETKDLYAPKPSGK	BLK, LYN
PCDH18	54510	NP_061908	Y743	CRVAESTYQHHPKRP	FES, FynTR
PCDH7	5099	NP_115833	Y901	KKSKQPLYSSIVTVE	BLK, HCK
PCGF2	7703	NP_009075	Y197	KYKVEVLYEDEPLKE	FES
PCM1	5108	NP_006188	Y1976	LPLKLTIYSEADLRK	BMX
PDCD4	27250	NP_055271	Y152	DDQENCVYETVVLPL	FER
PDGFRB	5159	NP_002600	Y778	YMAPYDNYVPSAPER	PTK2
PDK1	5163	NP_002601	Y357	VPLAGFGYGLPISRL	SRMS
PDLIM1	9124	NP_066272	Y144	ARVITNQYNNPAGLY	FES
PDLIM2	64236	NP_789847	Y172	PAADRLSYSGRPGSR	SYK
PDXDC1	23042	NP_055842	Y677	LESTEPIYVYKAQGA	FRK, HCK
PEAK1	79834	NP_079052	Y1348	AVYYTASYAKDPLNN	SYK
PEAR1	375033	NP_001073940	Y925	SLSSENPYATIRDLP	FRK, FynTR, YES1
PEBP1	5037	NP_002558	Y181	DDYVPKLYEQLSGK-	SRC
PECAM1	5175	NP_000433	Y713	KKDTETVYSEVRKAV	BLK, FRK, FYN, LYN
PFAS	5198	NP_036525	Y538	DPAGAIIYTSRFQLG	HCK, SRC
PFN2	5217	NP_444252	Y99	SQGGEPTYNVAVGRA	FER
PGM5	5239	NP_068800	Y358	KSMKVPVYETPAGWR	SRC
PHACTR2	9749	NP_001093634	Y397	PTNRTTLYSGTGLSV	SRC, YES1
PHB	5245	NP_002625	Y249	EAAEDIAYQLSRSRN	ABL2, FER
PHB2	11331	NP_009204	Y248	ALSKNPGYIKLRKIR	ABL2, FRK

PHF2	5253	NP_005383	Y677	PKPVRDEYEYVSDDG	FRK
PHF2	5253	NP_005383	Y904	KGSDDAPYSPTARVG	FGR, HCK
PHKB	5257	NP_000284	Y29	TKRSGSVYEPLKSIN	LYN
PHKG1	5260	NP_006204	Y338	EIVIRDPYALRPLRR	FGR, FRK
PI4K2A	55361	NP_060895	Y18	ERAQPPDYTFPSGSG	FGR
PI4KA	5297	NP_477352	Y1096	NRYAGEVYGMIRFSG	BLK
PIK3AP1	118788	NP_689522	Y419	GEEADAVYESMAHLS	HCK
PIK3AP1	118788	NP_689522	Y444	PGCDEDLYESMAAFV	FRK
PIK3C2B	5287	NP_002637	Y1541	DGNDPDPYVKIYLLP	HCK
PIK3CD	5293	NP_005017	Y524	RRGSGELYEHEKDLV	BMX
PIK3R1	5295	NP_852664	Y556	LKKQAAEYREIDKRM	BMX
PIK3R2	5296	NP_005018	Y464	SREYDQLYEEYTRTS	FRK
PIK3R3	8503	NP_003620	Y373	VQAEDLLYGKPDGAF	FER
PIKFYVE	200576	NP_055855	Y1773	RGADSAYYQVGQTGK	TNK1
PKP2	5318	NP_004563	Y86	RTSSVPEYVYNLHLV	BMX
PKP2	5318	NP_004563	Y130	WGRGTAQYSSQKSVE	HCK
PKP4	8502	NP_003619	Y275	AARAASPYSQRPASP	ABL2, FGR, PTK2, SYK, TNK1
PKP4	8502	NP_003619	Y306	PNGPTPQYQTTARVG	FRK, LYN
PKP4	8502	NP_003619	Y415	DLHITPIYEGRTYYS	BLK, YES1
PKP4	8502	NP_003619	Y478	ALNTTATYAEPYRPI	FRK
PLCG1	5335	NP_877963	Y210	QRSGDITYGQFAQLY	YES1
PLCG1	5335	NP_877963	Y472	KLAEGSAYEEVPTSM	FGR
PLCG2	5336	NP_002652	Y1197	LESEEELYSSCRQLR	BLK, HCK, YES1
PLD2	5338	NP_002654	Y415	ALGINSGYSKRALML	PTK2
PLEC	5339	NP_958782	Y3362	RARQEELYSELQARE	BMX, HCK
PLEKHA1	59338	NP_067635	Y181	RSQSHLPYFTPKPPQ	TNK1
PLEKHA5	54477	NP_061885	Y398	RPNTGPLYTEADRVI	FYN
PLEKHA5	54477	NP_061885	Y436	ETRGVISYQTLPRNM	FRK
PLEKHA6	22874	NP_055750	Y350	VSRRVPEYYGPYSSQ	BLK
PLEKHA6	22874	NP_055750	Y492	PPRSEDIYADPAAYV	ABL2, BLK
PLEKHA7	144100	NP_778228	Y656	GKSADDTYLQLKKDL	ABL2
PLEKHB1	58473	NP_067023	Y176	YSPYQDYYEVVPPNA	HCK
PLIN3	10226	NP_005808	Y95	QIASASEYAHRGLDK	BMX
PLK1	5347	NP_005021	Y217	TLCGTPNYIAPEVLS	ABL2
PLXNB1	5364	NP_002664	Y1708	DSVGEPLYMLFRGIK	SRC
PMPCB	9512	NP_004270	Y141	TSREQTVYYAKAFSK	BLK
POF1B	79983	NP_079197	Y49	PPEKNVVYERVRTYS	BLK
POLD3	10714	NP_006582	Y42	NQAKQMLYDYVERKR	ABL2, BLK
POLR2A	5430	NP_000928	Y145	KKRLTHVYDLCKGKN	FYN
POLR2A	5430	NP_000928	Y1895	YSPTSPTYSPTSPVY	PTK2
POLR2A	5430	NP_000928	Y1916	YSPTSPTYSPTSPKY	PTK2
POTEE	445582	NP_001077007	Y866	VTHTVPIYEGNALPH	BLK, FER, SRC
POTEF	728378	NP_001093241	Y866	VTHTVPIYEGNALPH	BLK, FER, SRC
PPIL4	85313	NP_624311	Y466	SKYQTDLYERERSKK	HCK, YES1
PPP1R12A	4659	NP_002471	Y766	SRTYDETYQRYRPVS	FES, HCK
PPP1R12B	4660	NP_002472	Y549	QTIAPSTYVSTYLKR	FRK, HCK, PTK2
PPP1R13B	23368	NP_056131	Y349	RVAAVGPYIQVPSAG	TNK1
PPP1R3B	79660	NP_078883	Y285	YEKLGPYY-------	FES, LYN, SRC, SRMS, SYK
PPP1R9B	84687	NP_115984	Y748	LRETQAQYQALERKY	FRK
PRC1	9055	NP_003972	Y464	QTETEMLYGSAPRTP	HCK

PRIC285	85441	NP_001032412	Y1304	LRILGLEYSLRVPPS	BMX
PRICKLE4	29964	NP_037529	Y136	ELLKPGEYGVFAARA	FGR, SRC
PRKAA1	5562	NP_996790	Y456	EWKVVNPYYLRVRRK	FynTR, YES1
PRKACB	5567	NP_891993	Y116	KHKATEQYYAMKILD	FRK
PRKAR1B	5575	NP_002726	Y312	RRSPNEEYVEVGRLG	FRK
PRKCA	5578	NP_002728	Y195	PNGLSDPYVKLKLIP	FRK
PRKCB	5579	NP_002729	Y195	PNGLSDPYVKLKLIP	FRK
PRKCD	5580	NP_006245	Y313	SSEPVGIYQGFEKKT	YES1
PRKCD	5580	NP_006245	Y374	ELKGRGEYFAIKALK	FGR
PRKCG	5582	NP_002730	Y195	PNGLSDPYVKLKLIP	FRK
PRKCQ	5588	NP_006248	Y90	SETTVELYSLAERCR	ABL2, FER, FynTR
PRKCZ	5590	NP_002735	Y417	TFCGTPNYIAPEILR	ABL2
PRMT1	3276	NP_001527	Y309	STSPESPYTHWKQTV	FGR, YES1
PRMT5	10419	NP_006100	Y297	NRPPPNAYELFAKGY	FGR
PRMT5	10419	NP_006100	Y342	SQYQQAIYKCLLDRV	SRMS
PRPF8	10594	NP_006436	Y2102	DDIKETGYTYILPKN	SRMS
PRRC2A	7916	NP_542417	Y1031	RGRGRGEYFARGRGF	FGR
PRRC2A	7916	NP_542417	Y1094	TRSEGSEYEEIPKRR	BMX, FGR
PRRC2C	23215	NP_055987	Y164	KETNDDNYGPGPSLR	LYN
PRUNE2	158471	NP_056040	Y1738	DPKSENIYDYLDSSE	SRC, YES1
PSMA6	5687	NP_002782	Y160	KCDPAGYYCGFKATA	FGR, TNK1, YES1
PSMA7	5688	NP_002783	Y153	QTDPSGTYHAWKANA	FGR, PTK2, YES1
PSMB5	5693	NP_002788	Y220	DLARRAIYQATYRDA	PTK2
PSMC4	5704	NP_006494	Y205	PPRGVLMYGPPGCGK	BLK, FynTR
PSMC5	5705	NP_002796	Y189	QPKGVLLYGPPGTGK	HCK, TNK1
PSMD11	5717	NP_002806	Y415	SKVVDSLYNKAKKLT	FER
PSMD9	5715	NP_002804	Y70	PRSDVDLYQVRTARH	ABL2, BLK, YES1
PSTPIP1	9051	NP_003969	Y345	PERNEGVYTAIAVQE	FER
PSTPIP2	9050	NP_077748	Y333	VDDYSLLYQ------	ABL2
PTEN	5728	NP_000305	Y178	QRRYVYYYSYLLKNH	BMX
PTGDS	5730	NP_000945	Y149	DFRMATLYSRTQTPR	FER
PTGES2	80142	NP_079348	Y222	EKEAQQVYGGKEART	HCK
PTK2	5747	NP_005598	Y419	SVSETDDYAEIIDEE	SYK
PTK2	5747	NP_005598	Y598	RYMEDSTYYKASKGK	BLK
PTPN12	5782	NP_002826	Y301	FEKQLQLYEIHGAQK	BMX
PTPN14	5784	NP_005392	Y445	SYRPTPDYETVMRQM	FES
PTPN18	26469	NP_055184	Y389	SAEEAPLYSKVTPRA	YES1
PTPN18	26469	NP_055184	Y426	SPAGSGAYEDVAGGA	FER, FGR
PTPN21	11099	NP_008970	Y569	VYRPPPPYPPPRPAN	TNK1
PTPRF	5792	NP_002831	Y224	YSAPANLYVRVRRVA	FES, SRC
PTPRT	11122	NP_573400	Y912	GYGFKEEYEALPEGQ	BLK
PVRL2	5819	NP_001036189	Y513	LDKINPIYDALSYSS	ABL2, BLK, SYK
PXN	5829	NP_001074324	Y40	YPTGNHTYQEIAVPP	FYN
PYGM	5837	NP_005600	Y733	GYNAQEYYDRIPELR	FRK
PZP	5858	NP_002855	Y700	AGAVGQGYYGAGLGV	FRK
QARS	5859	NP_005042	Y491	YGRLNLHYAVVSKRK	FER
RAD54L2	23132	NP_055921	Y982	DKRTSVPYTRPSYAQ	ABL2, FER
RALY	22913	NP_057951	Y109	KRAASAIYSGYIFDY	PTK2
RANBP3	8498	NP_015561	Y115	SDREDGNYCPPVKRE	LYN
RAP1A	5906	NP_002875	Y159	INVNEIFYDLVRQIN	FRK
RAP1B	5908	NP_056461	Y159	INVNEIFYDLVRQIN	FRK
RARS	5917	NP_002878	Y536	NTAAYLLYAFTRIRS	HCK, SRC, YES1

RASA1	5921	NP_002881	Y460	TVDGKEIYNTIRRKT	HCK
RB1	5925	NP_000312	Y790	PHIPRSPYKFPSSPL	PTK2
RB1	5925	NP_000312	Y805	RIPGGNIYISPLKSP	LYN
RBL1	5933	NP_002886	Y1004	TPRSALLYKFNGSPS	LYN
RBL2	5934	NP_005602	Y667	ITSPTTLYDRYSSPP	FER, FRK, FYN, YES1
RBM10	8241	NP_005667	Y732	PRGLVAAYSGESDSE	SYK
RBM15	64783	NP_073605	Y251	PLKIEAVYVSRRRSR	FGR, FRK, FYN, HCK
RBM15	64783	NP_073605	Y336	RERDYPFYERVRPAY	FER, FES
RBM15	64783	NP_073605	Y537	EHRYQQQYLQPLPLT	ABL2
RBM22	55696	NP_060517	Y156	KLARTTPYYKRNRPH	SRC
RBM3	5935	NP_006734	Y125	YDSRPGGYGYGYGRS	TNK1
RBM3	5935	NP_006734	Y143	NGRNQGGYDRYSGGN	FER, FRK
RBM3	5935	NP_006734	Y155	GGNYRDNYDN-----	FES, SRMS
RBM33	155435	NP_444271	Y835	AQQQQQLYAPPPPAE	TNK1
RBM4	5936	NP_002887	Y113	ECDIVKDYAFVHMER	FGR
RBM4B	83759	NP_113680	Y113	ECDIVKDYAFVHMER	FGR
RBM5	10181	NP_005769	Y76	SDRSEDGYHSDGDYG	PTK2
RBM5	10181	NP_005769	Y620	KRGLVAAYSGDSDNE	SYK
RBM5	10181	NP_005769	Y733	FDAGTVNYEQPTKDG	ABL2
RBMX	27316	NP_002130	Y320	YSSSRDGYGGSRDSY	PTK2
RBMX	27316	NP_002130	Y327	YGGSRDSYSSSRSDL	PTK2
RCAN3	11123	NP_038469	Y120	NGQKLKLYFAQVQMS	SYK
RCL1	10171	NP_005763	Y74	NQTGTTLYYQPGLLY	ABL2
RCOR3	55758	NP_001129695	Y185	IASLVKYYYSWKKTR	SRC, YES1
REPS2	9185	NP_004717	Y558	PAKKDVLYSQPPSKP	ABL2, SYK
RET	5979	NP_066124	Y952	VTLGGNPYPGIPPER	TNK1
REV3L	5980	NP_002903	Y128	HFMKIYLYNPTMVKR	SRMS
REXO2	25996	NP_056338	Y122	ITLQQAEYEFLSFVR	YES1
RFC1	5981	NP_002904	Y168	TPTSVLDYFGTGSVQ	SRC
RFFL	117584	NP_476519	Y29	QGARMQAYSNPGYSS	PTK2
RFT1	91869	NP_443091	Y55	NVRLTLLYSTTLFLA	SYK
RFX4	5992	NP_998759	Y129	RGQSKYHYYGIAVKE	SYK
RGS14	10636	NP_006471	Y122	AQEARNIYQEFLSSQ	LYN
RGS19	10287	NP_005864	Y143	DEKARLIYEDYVSIL	FES, SYK
RIF1	55183	NP_060621	Y404	HKGASSPYGAPGTPR	FER
RIN1	9610	NP_004283	Y36	KPAQDPLYDVPNASG	ABL2, YES1
RIPK1	8737	NP_003795	Y384	KLQDEANYHLYGSRM	SRMS
RIPK2	8767	NP_003812	Y474	DLIMKEDYELVSTKP	FES
RLN2	6019	NP_604390	Y164	SRKKRQLYSALANKC	FES
RNASEH2A	10535	NP_006388	Y172	KAKADALYPVVSAAS	TNK1
RNF148	378925	NP_932351	Y130	GANGVIIYNYQGTGS	FER, FES, HCK, YES1
RNF181	51255	NP_057578	Y152	ENLHGAMYT------	SRMS
ROBO1	6091	NP_002932	Y932	RNGLTSTYAGIRKVP	FRK, LYN
ROBO1	6091	NP_002932	Y1073	PSGQPTPYATTQLIQ	FGR
ROR1	4919	NP_005003	Y645	REIYSADYYRVQSKS	FES
ROR1	4919	NP_005003	Y836	AAFPAAHYQPTGPPR	PTK2, TNK1
ROS1	6098	NP_002935	Y2334	GKPEGLNYACLTHSG	ABL2
RP1	6101	NP_006260	Y1525	KRMNGIIYEIISKRL	BLK
RPL10A	4736	NP_009035	Y11	KVSRDTLYEAVREVL	BLK, BMX, FynTR
RPL15	6138	NP_002939	Y62	AKQGYVIYRIRVRRG	ABL2
RPL23A	6147	NP_000975	Y144	YVRLAPDYDALDVAN	SYK
RPL27A	6157	NP_000981	Y52	FDKYHPGYFGKVGMK	SYK

RPL31	6160	NP_001092047	Y103	EDSPNKLYTLVTYVP	FER, FRK
RPL35	11224	NP_009140	Y78	KFYKGKKYKPLDLRP	SRMS
RPL5	6125	NP_000960	Y44	VIQDKNKYNTPKYRM	SRMS
RPL8	6132	NP_000964	Y133	LARASGNYATVISHN	FES, TNK1
RPL9	6133	NP_000652	Y180	RKFLDGIYVSEKGTV	FYN
RPS10	6204	NP_001005	Y12	KKNRIAIYELLFKEG	SRC
RPS13	6207	NP_001008	Y38	DDVKEQIYKLAKKGL	FynTR
RPS16	6217	NP_001011	Y82	GGHVAQIYAIRQSIS	ABL2, BMX, HCK
RPS6	6194	NP_001001	Y209	NKEEAAEYAKLLAKR	BMX, SYK
RPS6KA4	8986	NP_003933	Y44	KVLGTGAYGKVFLVR	BLK, FYN
RPS6KA4	8986	NP_003933	Y342	FTRLEPVYSPPGSPP	FER
RPS6KA6	27330	NP_055311	Y231	VDQEKKAYSFCGTVE	FER
RPTOR	57521	NP_065812	Y723	RLRSVSSYGNIRAVA	PTK2
RRP7A	27341	NP_056518	Y23	RIPSPLGYAAIPIKF	TNK1
RSF1	51773	NP_057662	Y1281	RGRSTDEYSEADEEE	SYK
RSRC2	65117	NP_075388	Y317	TGIAVPSYYNPAAVN	PTK2
RUVBL2	10856	NP_006657	Y172	TTEMETIYDLGTKMI	YES1
RYBP	23429	NP_036366	Y70	AQQVAQQYATPPPPK	ABL2
S100A11	6282	NP_005611	Y30	KYAGKDGYNYTLSKT	SRMS
SACS	26278	NP_055178	Y593	YTKTVLNYLQSSGKQ	TNK1
SACS	26278	NP_055178	Y3463	KIHLKELYEVIGCVP	FER, FYN, HCK
SAGE1	55511	NP_061136	Y791	FAVGTKNYSVSAGDP	FER
SAMD10	140700	NP_542188	Y138	CPHNYLVYVEAFSQH	FYN
SAMSN1	64092	NP_071419	Y160	RLDDDGPYSGPFCGR	FynTR
SCAF8	22828	NP_055707	Y1237	VQNDPELYEKLTSSN	ABL2
SCAMP3	10067	NP_005689	Y53	TREPPPAYEPPAPAP	TNK1
SCAMP3	10067	NP_005689	Y86	EPKNYGSYSTQASAA	TNK1
SCARF2	91179	NP_699165	Y628	EGPGGALYARVARRE	FER, HCK, TNK1
SCEL	8796	NP_659001	Y295	KSLESLIYMSTRTDK	HCK
SCG2	7857	NP_003460	Y349	PLDSQSIYQLIEISR	FRK
SCML2	10389	NP_006080	Y506	SPQQTVPYVVPLSPK	ABL2
SCRIB	23513	NP_874365	Y834	PLRPEDDYSPRERRG	FGR
SDC2	6383	NP_002989	Y191	RKPSSAAYQKAPTKE	FER
SEC16A	9919	NP_055681	Y1209	YDGAASAYAQNYRYP	SYK
SEC16A	9919	NP_055681	Y1243	QGYPEGYYSSKSGWS	FRK
SEC24B	10427	NP_006314	Y117	QPGAQQLYSRGPPAP	TNK1
SEC24C	9632	NP_004913	Y296	ARGPQSNYGGPYPAA	PTK2, TNK1
SEC31A	22872	NP_055748	Y804	HESPKIPYEKQQLPK	FER
SERBP1	26135	NP_001018077	Y246	QKQISYNYSDLDQSN	ABL2
SETD1A	9739	NP_055527	Y370	DANYPAYYESWNRYQ	YES1
SF3B1	23451	NP_036565	Y421	VLPPPAGYVPIRTPA	PTK2, TNK1
SFPQ	6421	NP_005057	Y698	YGRGREEYEGPNKKP	FES, FRK
SGMS2	166929	NP_689834	Y24	PSDPTNTYARPAEPV	FRK
SGTA	6449	NP_003012	Y158	AICIDPAYSKAYGRM	FYN
SH2B1	25970	NP_056318	Y55	YLASHPQYAGPGAEA	TNK1
SH2B2	10603	NP_066189	Y672	ARAVENQYSFY----	PTK2
SH2D2A	9047	NP_003966	Y39	RSCQNLGYTAASPQA	TNK1
SH2D3C	10044	NP_733745	Y316	EQSGAIIYCPVNRTF	FER
SH3BP2	6452	NP_003014	Y448	GDDSDEDYEKVPLPN	FER, FYN
SH3GLB2	56904	NP_064530	Y77	ARVEEFLYEKLDRKV	YES1
SH3PXD2B	285590	NP_001017995	Y25	RRVPNKHYVYIIRVT	FES
SHANK2	22941	NP_036441	Y989	GQMPENPYSEVGKIA	SRC

SHANK2	22941	NP_036441	Y1201	GTAGPGNYVHPLTGR	TNK1
SHANK3	85358	NP_001073889	Y887	GAGAAGLYEPGAALG	LYN
SHC2	25759	NP_036567	Y339	DSLEHNYYNSIPGKE	SRC
SHC4	399694	NP_976224	Y424	NSKCSSVYENCLEQS	LYN
SHCBP1	79801	NP_079021	Y217	DEEEEDEYDYFVRCV	BLK
SHISA5	51246	NP_057563	Y221	AGGAAAPYPASQPPY	TNK1
SIGLEC7	27036	NP_055200	Y460	QEATNNEYSEIKIPK	ABL2, FRK
SIN3A	25942	NP_056292	Y565	YRALPKSYQQPKCTG	FRK
SIN3B	23309	NP_056075	Y1155	VTRYRVQYSRRPASP	LYN
SIPA1L3	23094	NP_055888	Y1141	VSFPETPYTVSPAGA	FGR
SIPA1L3	23094	NP_055888	Y1169	STPGSATYVRYKPSP	HCK, PTK2
SIRPA	140885	NP_542970	Y496	PEPSFSEYASVQVPR	FRK
SIT1	27240	NP_055265	Y188	ASFPDQAYANSQPAA	TNK1
SKAP1	8631	NP_003717	Y295	TRRKGVDYASYYQGL	FYN
SKAP2	8935	NP_003921	Y261	QPIDDEIYEELPEEE	BMX
SLAMF7	57823	NP_067004	Y304	EDPANTVYSTVEIPK	BLK
SLC12A2	6558	NP_001037	Y275	TRDAVVTYTAESKGV	FYN
SLC12A4	6560	NP_005063	Y17	DGPRRGDYDNLEGLS	FES
SLC20A2	6575	NP_006740	Y377	KPAQESNYRLLRRNN	ABL2
SLC22A20	440044	NP_001004326	Y393	VATATMIYVGRRATV	HCK
SLC25A1	6576	NP_005975	Y276	KEGLKAFYKGTVPRL	SRMS
SLC25A24	29957	NP_037518	Y328	TGQYSGIYDCAKKIL	ABL2
SLC25A3	5250	NP_005879	Y196	RIQTQPGYANTLRDA	TNK1
SLC25A5	292	NP_001143	Y191	IIIYRAAYFGIYDTA	PTK2
SLC26A3	1811	NP_000102	Y520	NIGRTNIYKNKKDYY	LYN
SLC45A1	50651	NP_001073866	Y628	ATLSRNLYVVLSLCI	BMX
SLC4A1	6521	NP_000333	Y904	EEEGRDEYDEVAMPV	FRK
SLC4A7	9497	NP_003606	Y87	DGRESPSYDTPSQRV	PTK2
SLC8A2	6543	NP_055878	Y255	KRLLFYKYVYKRYRT	FES
SLITRK2	84631	NP_115928	Y756	PREPELLYQNIAERV	LYN
SLITRK5	26050	NP_056382	Y749	KTPAGHVYEYIPHPL	FYN
SLITRK6	84189	NP_115605	Y805	KLMETLMYSRPRKVL	BLK
SMARCE1	6605	NP_003070	Y142	NSPAYLAYINAKSRA	TNK1
SMC3	9126	NP_005436	Y669	GALTGGYYDTRKSRL	BLK, FYN
SMC4	10051	NP_005487	Y175	IDKEGDDYEVIPNSN	ABL2, FYN
SMG8	55181	NP_060619	Y135	TEAGSQDYSLLQAYY	ABL2, FGR
SNAP29	9342	NP_004773	Y68	SRSLALMYESEKVGV	YES1
SNRNP200	23020	NP_054733	Y1754	NKQDAVDYLTWTFLY	ABL2
SNTB1	6641	NP_066301	Y483	KTIIQSPYEKLKMSS	FRK
SNX9	51429	NP_057308	Y177	DGPKSSSYFKDSESA	PTK2
SNX9	51429	NP_057308	Y219	AKPGTEQYLLAKQLA	ABL2
SON	6651	NP_620305	Y2192	KKEADSVYGEWVPVE	HCK
SORBS1	10580	NP_001030126	Y536	RAEPKSIYEYQPGKS	SYK
SOWAHA	134548	NP_787069	Y425	SGRRAYQYLRPGSSY	ABL2, PTK2
SPATA13	221178	NP_694568	Y404	TTQEHGDYSNIKAAY	SYK
SPATA5	166378	NP_660208	Y393	APRGVLLYGPPGTGK	FynTR, HCK, LYN, TNK1
SPATS2L	26010	NP_056350	Y15	VNVKEKIYAVRSVVP	FER, FYN
SPNS1	83985	NP_114427	Y167	VGVGEASYSTIAPTL	FER, SRMS
SPTBN1	6711	NP_003119	Y796	VAEEIANYRPTLDTL	SRMS
SPTBN1	6711	NP_003119	Y1680	QSKVDKLYAGLKDLA	FRK, SRC, YES1
SPTLC1	10558	NP_006406	Y82	KDHPALNYNIVSGPP	FES, SRC

SRC	6714	NP_005408	Y419	RLIEDNEYTARQGAK	FGR, FRK
SRCIN1	80725	NP_079524	Y264	IYRKEPLYAAFPGSH	BLK, FYN, SRC
SRRM4	84530	NP_919262	Y457	RSRRSPSYSRYSPSR	PTK2, TNK1
SRRT	51593	NP_056992	Y836	YGAGRGNYDAFRGQG	PTK2
SRSF1	6426	NP_008855	Y149	REAGDVCYADVYRDG	BLK
SRSF1	6426	NP_008855	Y170	VRKEDMTYAVRKLDN	BLK, SRC
SRSF1	6426	NP_008855	Y202	DGPRSPSYGRSRSRS	PTK2
SRSF10	10772	NP_473357	Y138	SRSFDYNYRRSYSPR	ABL2
SRSF4	6429	NP_005617	Y53	RDADDAVYELNGKDL	BMX
SRSF6	6431	NP_006266	Y53	RDADDAVYELNGKEL	BMX
SRSF8	10929	NP_115285	Y175	SRYSRSPYSRSRYRE	PTK2
SRSF9	8683	NP_003760	Y139	REAGDVCYADVQKDG	BLK
SRSF9	8683	NP_003760	Y221	YFSPFRPY-------	SRMS
SS18	6760	NP_001007560	Y416	DQGQYGNYQQ-----	ABL2, SRMS
SSBP1	6742	NP_003134	Y99	PGLRDVAYQYVKKGS	BLK
SSBP2	23635	NP_036578	Y192	VPLGPQNYGGAMRPP	TNK1
SSBP3	23648	NP_663768	Y23	AREKLALYVYEYLLH	BMX
SSBP3	23648	NP_663768	Y25	EKLALYVYEYLLHVG	BMX
ST3GAL5	8869	NP_003887	Y35	AMPSEYTYVKLRSDC	FYN, PTK2
ST5	6764	NP_005409	Y488	STLEENAYEDIVGDL	FER
ST5	6764	NP_005409	Y501	DLPKENPYEDVDLKS	FES, SRC
STAG2	10735	NP_001036214	Y433	VAAGEFLYKKLFSRR	SRC
STAM2	10254	NP_005834	Y374	EAPVYSVYSKLHPPA	FER, FYN, HCK, SRC
STAP2	55620	NP_060190	Y322	GDGPAVDYENQDVAS	SYK
STAT5A	6776	NP_003143	Y694	LAKAVDGYVKPQIKQ	ABL2
STAT5B	6777	NP_036580	Y679	DRPKDEVYSKYYTPV	BLK, HCK
STAT5B	6777	NP_036580	Y699	TAKAVDGYVKPQIKQ	ABL2
STAT5B	6777	NP_036580	Y740	AVCPQAHYNMYPQNP	PTK2
STIM2	57620	NP_065911	Y704	PPLSLEIYQTLSPRK	BLK, BMX, SRMS
STIP1	10963	NP_006810	Y41	DPHNHVLYSNRSAAY	BLK
STK35	140901	NP_543026	Y217	RGSYGVVYEAVAGRS	BLK, FER
STK39	27347	NP_037365	Y273	LATGAAPYHKYPPMK	TNK1
STON1	11037	NP_006864	Y628	SAKYESAYQAVVWKI	FER
STT3B	201595	NP_849193	Y511	KRNQGNLYDKAGKVR	BMX, FER, FynTR, HCK, LYN
SUPT16H	11198	NP_009123	Y701	GDKVDILYNNIKHAL	BLK, LYN
SUPT5H	6829	NP_003160	Y140	REEELGEYYMKKYAK	SYK
SV2A	9900	NP_055664	Y41	LDRVQDEYSRRSYSR	ABL2, BLK
SYK	6850	NP_003168	Y74	ERELNGTYAIAGGRT	BLK, FYN
SYK	6850	NP_003168	Y630	ELRLRNYYYDVVN--	FES
SYK	6850	NP_003168	Y631	LRLRNYYYDVVN---	BMX, SRMS
SYNCRIP	10492	NP_006363	Y373	RVKKLKDYAFIHFDE	SYK
SYNCRIP	10492	NP_006363	Y485	RGGYEDPYYGYEDFQ	PTK2
SYNE1	23345	NP_892006	Y6742	LKSSLNEYQPKLYQV	BMX
SYNE1	23345	NP_892006	Y6747	NEYQPKLYQVLDDGK	YES1
SYNGR2	9144	NP_004701	Y203	PGASVDNYQQPPFTQ	ABL2, PTK2, TNK1
SYNPO	11346	NP_009217	Y565	EPTKQPPYQLRPSLF	FynTR
SYNPO	11346	NP_009217	Y624	SRSSPGLYTSPGQDS	FRK
SYT11	23208	NP_689493	Y337	PYVKVNVYYGRKRIA	FGR, FRK, HCK
SYT11	23208	NP_689493	Y338	YVKVNVYYGRKRIAK	BLK
SYT14	255928	NP_694994	Y454	ANRPPNTYVKLTLLN	FRK
SZT2	23334	NP_056099	Y1228	SRSQEPIYSEEASGP	FynTR, HCK, YES1

TACC2	10579	NP_996744	Y164	ERDSSTPYQEIAAVP	FGR, FRK, HCK
TAF15	8148	NP_631961	Y83	SSYSQQPYNNQGQQQ	PTK2
TAF15	8148	NP_631961	Y463	GGDRGGGYGGDRGGG	TNK1
TAF6	6878	NP_005632	Y253	IATDPGLYQMLPRFS	BLK
TAL1	6886	NP_003180	Y138	APGRALLYSLSQPLA	BMX
TANC2	26115	NP_079461	Y1964	QAYQDNLYRQLSRDS	ABL2
TAOK2	9344	NP_057235	Y426	LPGSDNLYDDPYQPE	FER
TAP2	6891	NP_000535	Y693	LQEGQDLYSRLVQQR	ABL2
TBC1D10A	83874	NP_114143	Y328	VKACQGQYETIERLR	LYN
TBC1D19	55296	NP_060787	Y25	KLKGSNLYSQLERQA	ABL2, FRK, FynTR
TBC1D5	9779	NP_001127853	Y2	------MYHSLSETR	SRMS
TBP	6908	NP_003185	Y329	YEAFENIYPILKGFR	SRC
TCF12	6938	NP_996919	Y157	GKPGTAYYSFSATSS	FER
TCF3	6929	NP_003191	Y153	TSQYYPSYSGSSRRR	PTK2
TCHP	84260	NP_115676	Y140	LIAEQLLYEHWKKNN	YES1
TCP1	6950	NP_110379	Y181	DAVLAIKYTDIRGQP	FES
TDP2	51567	NP_057698	Y158	QEVIPPYYSYLKKRS	HCK, YES1
TEK	7010	NP_000450	Y1048	GMTCAELYEKLPQGY	FynTR
TES	26136	NP_056456	Y111	VSINTVTYEWAPPVQ	FER
TES	26136	NP_056456	Y251	KEGDPAIYAERAGYD	SRC
TEX10	54881	NP_060216	Y623	QRLVQLVYFLPSLPA	YES1
TEX2	55852	NP_060939	Y76	LEAKEDLYLEPQVGH	ABL2
TEX2	55852	NP_060939	Y299	RRLSEVIYEPFQLLS	BMX
TFG	10342	NP_006061	Y392	RPPFGQGYTQPGPGY	TNK1
TGM2	7052	NP_004604	Y219	SRRSSPVYVGRVVSG	HCK, YES1
TGM2	7052	NP_004604	Y369	QEKSEGTYCCGPVPV	YES1
TJP1	7082	NP_003248	Y833	PGSEYSMYSTDSRHT	SYK
TJP1	7082	NP_003248	Y1435	PPPLPSQYAQPSQPV	ABL2, TNK1
TJP2	9414	NP_004808	Y249	GRSIDQDYERAYHRA	FGR, FYN
TKT	7086	NP_001055	Y275	EQIIQEIYSQIQSKK	ABL2, BLK
TLN1	7094	NP_006280	Y570	GDPAETDYTAVGCAV	FGR, FYN
TLN1	7094	NP_006280	Y1116	VAQGNENYAGIAARD	TNK1
TLR4	7099	NP_612564	Y674	YGRGENIYDAFVIYS	BLK, FynTR, HCK, SRC
TM9SF4	9777	NP_055557	Y312	LRKDIANYNKEDDIE	SYK
TM9SF4	9777	NP_055557	Y636	YMFVRKIYAAVKID-	FYN
TMED7	51014	NP_861974	Y50	DNAKQCFYEDIAQGT	FER
TMEM200A	114801	NP_443145	Y164	IIHMRDIYSTVIDIH	YES1
TMEM201	199953	NP_001124396	Y153	YDEEVEVYRHHLEQM	BMX
TMEM57	55219	NP_060672	Y659	LDPNASVYQPLKK--	BMX, LYN
TMPRSS11F	389208	NP_997290	Y151	KKIEKALYQSLKTKQ	HCK
TMTC4	84899	NP_116202	Y607	RKYPDCYYNLGRLYA	FGR
TNK2	10188	NP_001010938	Y937	KKVSSTHYYLLPERP	FER
TNKS1BP1	85456	NP_203754	Y855	KRDSQGTYSSRDAEL	FYN
TNRC6A	27327	NP_055309	Y1251	HSLNIGDYNRTVGKG	FES
TNRC6B	23112	NP_055903	Y1081	GTTDSGPYFEKGGSH	FGR
TNRC6B	23112	NP_055903	Y1389	IDPESDPYVTPGSVL	HCK, PTK2
TNS1	7145	NP_072174	Y366	GPLDGSLYAKVKKKD	BLK, FYN, FynTR
TNS1	7145	NP_072174	Y796	PGRSYSPYDYQPCLA	FER, FGR, HCK
TNS1	7145	NP_072174	Y1254	CRHPAGVYQVSGLHN	FES
TNS1	7145	NP_072174	Y1326	RHVAYGGYSTPEDRR	PTK2
TNS1	7145	NP_072174	Y1345	RQSSASGYQAPSTPS	PTK2

TNS1	7145	NP_072174	Y1404	RAGSLPNYATINGKV	TNK1
TNS3	64759	NP_073585	Y354	GPVDGSLYAKVRKKS	FER, FynTR, HCK
TNS3	64759	NP_073585	Y584	QAYGQSSYSTQTWVR	FRK
TNS3	64759	NP_073585	Y855	PVSPETPYVKTALRH	SRC
TNS3	64759	NP_073585	Y1256	KGCSNEPYFGSLTAL	TNK1
TOM1L1	10040	NP_005477	Y460	AVTTEAIYEEIDAHQ	FynTR
TOM1L2	146691	NP_001076437	Y200	SSPPPAPYSAPQAPA	FGR, PTK2
TOMM34	10953	NP_006800	Y54	PEEESVLYSNRAACH	ABL2, BLK
TOP2B	7155	NP_001059	Y1604	RARKEVKYFAESDEE	SYK
TOPORS	10210	NP_005793	Y193	RERNASVYSPSGPVN	BMX, FER
TP53BP2	7159	NP_001026855	Y356	RVAAVGPYIQSSTMP	PTK2, TNK1
TP53RK	112858	NP_291028	Y201	EDKGVDLYVLEKAFL	FYN
TPD52L2	7165	NP_955392	Y106	DVQVSSAYVKTSEKL	PTK2
TPPP3	51673	NP_057224	Y170	AYKNAGTYDAKVKK-	FES, LYN, SYK
TRA2A	29896	NP_037425	Y264	RRRSPSPYYSRYRSR	FGR, FRK, HCK, PTK2
TRA2A	29896	NP_037425	Y265	RRSPSPYYSRYRSRS	FES, LYN, PTK2, YES1
TRA2B	6434	NP_004584	Y260	AQDRDQIYRRRSPSP	ABL2
TRA2B	6434	NP_004584	Y268	RRRSPSPYYSRGGYR	FGR, FRK, PTK2
TRA2B	6434	NP_004584	Y269	RRSPSPYYSRGGYRS	PTK2, TNK1
TRAF3IP3	80342	NP_079504	Y234	RSLENQLYTCTQKYS	HCK
TRAP1	10131	NP_057376	Y366	LGSSVALYSRKVLIQ	BMX, SYK
TRAT1	50852	NP_057472	Y79	EPMDENCYEQMKARP	YES1
TRIM28	10155	NP_005753	Y369	LLSKKLIYFQLHRAL	SYK
TRIM58	25893	NP_056246	Y55	DGAQGGVYACPQCRG	ABL2, FGR
TRIOBP	11078	NP_001034230	Y2325	SGKYQDVYVELSHIK	BMX
TRIP11	9321	NP_004230	Y452	LLKLNNEYEVIKSTA	BLK
TRIP6	7205	NP_003293	Y157	GGPTPASYTTASTPA	TNK1
TROAP	10024	NP_005471	Y172	QGVRASAYLAPRTPT	PTK2
TSC1	7248	NP_000359	Y312	GCATSTPYSTSRLML	PTK2
TSC2	7249	NP_000539	Y1760	RICEEAAYSNPSLPL	FER
TSG101	7251	NP_006283	Y390	TAGLSDLY-------	LYN
TSNAX	7257	NP_005990	Y237	FIGNTGPYEVSKKLY	FER
TSPAN3	10099	NP_005715	Y243	RRSRDPAYELLITGG	BLK
TSPAN8	7103	NP_004607	Y122	RIVNETLYENTKLLS	FES
TSSK2	23617	NP_443732	Y23	INLGKGSYAKVKSAY	FER
TTC21B	79809	NP_079029	Y160	ITRGKEPYTKKALKY	HCK
TTN	7273	NP_596870	Y7941	DAGQYNCYIENASGK	LYN
TTN	7273	NP_596870	Y11409	PIEDVTIYEKESASF	FYN
TTN	7273	NP_596870	Y16787	GLKEGDTYEYRVSAV	BLK, FGR, FYN, LYN
TTN	7273	NP_596870	Y17151	ILGYIVEYQKVGDEE	SYK
TTN	7273	NP_596870	Y18171	RCVENQIYEFRVQTK	FGR, YES1
TUBA1A	7846	NP_006000	Y451	GEEEGEEY-------	BMX
TUBA1B	10376	NP_006073	Y451	GEEEGEEY-------	BMX
TUBB	203068	NP_821133	Y208	CIDNEALYDICFRTL	FYN, FynTR
TUBB1	81027	NP_110400	Y208	CIDNEALYDICFRTL	FYN, FynTR
TUBB2B	347733	NP_821080	Y208	CIDNEALYDICFRTL	FYN, FynTR
TUBB3	10381	NP_006077	Y208	CIDNEALYDICFRTL	FYN, FynTR
TUBB4A	10382	NP_006078	Y208	CIDNEALYDICFRTL	FYN, FynTR
TUBB4B	10383	NP_006079	Y208	CIDNEALYDICFRTL	FYN, FynTR
TUBB6	84617	NP_115914	Y208	CIDNEALYDICFRTL	FYN, FynTR

TUBGCP3	10426	NP_006313	Y133	AHSTPYYYARPQTLP	FRK, TNK1
TWF1	5756	NP_002813	Y343	ELTADFLYEEVHPKQ	SRC
TWF2	11344	NP_009215	Y309	ELTAEFLYDEVHPKQ	SRC
TWIST1	7291	NP_000465	Y103	GGGSPQSYEELQTQR	FRK
TXNL4B	54957	NP_060323	Y146	IPKYDLLYQDI----	BLK, FER
TXNRD1	7296	NP_001087240	Y163	DLPKSYDYDLIIIGG	FES
TYRO3	7301	NP_006284	Y685	RKIYSGDYYRQGCAS	FES
TYROBP	7305	NP_003323	Y91	ITETESPYQELQGQR	FRK
TYROBP	7305	NP_003323	Y112	LNTQRPYYK------	SRMS
UBA1	7317	NP_003325	Y55	ADIDEGLYSRQLYVL	ABL2, BMX
UBE2J1	51465	NP_057105	Y5	---METRYNLKSPAV	SRMS
UBE2O	63893	NP_071349	Y643	EEEDVSVYDIADHPD	BMX
UBXN6	80700	NP_079517	Y336	RGLRKYNYTLLRVRL	SRMS
UGGT1	56886	NP_064505	Y741	AVANSMNYLTKKGMS	PTK2
UGGT2	55757	NP_064506	Y623	GPLPQALYNGEPFKH	YES1
UNC13A	23025	NP_001073890	Y788	KTGSSDPYVTVQVGK	FGR, FRK, HCK, YES1
UPF1	5976	NP_002902	Y113	EEDEEDTYYTKDLPI	FRK
UPF1	5976	NP_002902	Y1101	ALSQDSTYQGERAYQ	FynTR, HCK
UPF3B	65109	NP_542199	Y442	DRPAMQLYQPGARSR	LYN
URB1	9875	NP_055640	Y180	DLNKKTLYTLVTKRD	SRC
USP10	9100	NP_005144	Y54	KLPDGQEYQRIEFGV	FRK
USP15	9958	NP_006304	Y247	TAYKNYDYSEPGRNN	ABL2, SYK
USP24	23358	NP_056121	Y2024	VYDQTNPYTDVRRRY	FER, FGR
USP4	7375	NP_003354	Y192	NKYMSNTYEQLSKLD	YES1
USP54	159195	NP_689799	Y1529	YTGRTLNYQSLPHRS	LYN
USP6NL	9712	NP_001073960	Y727	VDTGAGGYSGNSGSP	PTK2
USP6NL	9712	NP_001073960	Y821	ASPSGYPYSGPPPPA	PTK2, TNK1
UTP11L	51118	NP_057121	Y50	DYRKKQEYLKALRKK	SYK
UTP6	55813	NP_060898	Y260	KDLQKEIYDDLQALH	FER
UTRN	7402	NP_009055	Y1757	LIAQEPLYQCLVTTE	HCK, YES1
UTRN	7402	NP_009055	Y2360	RKYEARLYILQQARR	FGR
VAV2	7410	NP_001127870	Y172	EDGGDDIYEDIIKVE	BLK
VBP1	7411	NP_003363	Y112	FLLADNLYCKASVPP	FER
VCL	7414	NP_054706	Y100	QMLQSDPYSVPARDY	ABL2, PTK2
VCL	7414	NP_054706	Y1133	WVRKTPWYQ------	SYK
VCP	7415	NP_009057	Y173	VETDPSPYCIVAPDT	HCK
VCP	7415	NP_009057	Y244	PPRGILLYGPPGTGK	FER
VCP	7415	NP_009057	Y805	EDNDDDLYG------	FYN, FynTR, HCK, SRC
VDAC1	7416	NP_003365	Y195	TEFGGSIYQKVNKKL	ABL2
VIM	7431	NP_003371	Y61	ASSPGGVYATRSSAV	LYN
VPS35	55737	NP_060676	Y791	PESEGPIYEGLIL--	BLK, FynTR, YES1
VPS52	6293	NP_072047	Y256	FRKPMTNYQIPQTAL	FES
VRTN	55237	NP_060698	Y436	PGISRSTYYNWRRKA	PTK2
WAC	51322	NP_057712	Y25	RRGDSQPYQALKYSS	ABL2, FGR, HCK
WAPAL	23063	NP_055860	Y633	RREDKELYTVVQHVK	BMX, FYN, HCK
WASF1	8936	NP_003922	Y125	PIPLQETYDVCEQPP	FynTR, HCK
WASF3	10810	NP_006637	Y248	HASDVTDYSYPATPN	FER
WDR61	80349	NP_079510	Y301	DDQEIHIYDCPI---	SRMS
WEE1	7465	NP_003381	Y132	VKSPAAPYFLGSSFS	SRC
WRNIP1	56897	NP_064520	Y500	DRAGEEHYNCISALH	FER

WRNIP1	56897	NP_064520	Y534	EGGEDPLYVARRLVR	BMX, FRK, HCK
XPO7	23039	NP_055839	Y624	LNDLSIGYSSVRKLV	LYN
YES1	7525	NP_005424	Y141	IATGKNGYIPSNYVA	PTK2
YES1	7525	NP_005424	Y146	NGYIPSNYVAPADSI	FRK
YES1	7525	NP_005424	Y426	RLIEDNEYTARQGAK	FGR, FRK
YIPF5	81555	NP_110426	Y67	QPYTGQIYQPTQAYT	BMX, HCK
YLPM1	56252	NP_062535	Y1695	RERDREPYFDRQSNV	FGR
YTHDF1	54915	NP_060268	Y317	PALAQPQYQSPQQPP	FRK, HCK, LYN, PTK2
YTHDF3	253943	NP_689971	Y151	QSTQSSAYSSSYGYP	PTK2
ZAP70	7535	NP_001070	Y178	EEAERKLYSGAQTDG	FRK
ZAP70	7535	NP_001070	Y492	ALGADDSYYTARSAG	FRK
ZAP70	7535	NP_001070	Y493	LGADDSYYTARSAGK	BMX, FER, FRK, HCK
ZAP70	7535	NP_001070	Y598	QRMRACYYSLASKVE	SRC
ZBTB33	10009	NP_006768	Y442	ANIGEDTYDIVIPVK	FYN
ZC3H13	23091	NP_055885	Y359	KRTPSPSYQRTLTPP	PTK2, TNK1
ZC3HAV1	56829	NP_064504	Y642	SSYLESLYQSCPRGV	HCK, LYN
ZCCHC3	85364	NP_149080	Y201	GMDPSDIYAVIQIPG	FYN
ZDHHC17	23390	NP_056151	Y70	KATQYGIYERCRELV	FynTR, LYN
ZDHHC5	25921	NP_056272	Y91	DDFRAPLYKTVEIKG	SRC
ZFC3H1	196441	NP_659419	Y974	RRLQKLEYEYALKIQ	BMX
ZFR	51663	NP_057191	Y186	PSVAETYYQTAPKAG	FRK
ZNF185	7739	NP_009081	Y408	PKGALADYEGKDVAT	SYK
ZNF24	7572	NP_008896	Y279	IHSGEKPYGCVECGK	FynTR
ZNF24	7572	NP_008896	Y335	IHTGEKPYECVQCGK	FynTR
ZNF264	9422	NP_003408	Y343	VHSGENPYECLECGK	ABL2, FynTR
ZNF264	9422	NP_003408	Y483	IHTGEKPYECVECGK	FynTR
ZNF264	9422	NP_003408	Y511	IHSGEKPYECVECGK	FynTR
ZNF330	27309	NP_055302	Y275	HDEEEDEYEAEDDEE	SYK
ZNF433	163059	NP_001073880	Y134	TGHKAYEYQEYGQKP	BMX
ZNF433	163059	NP_001073880	Y142	QEYGQKPYKCKYCKK	HCK
ZNF460	10794	NP_006626	Y420	IHTGEKPYECLQCGK	FynTR
ZNF460	10794	NP_006626	Y476	IHTGEKPYECVECGK	FynTR
ZNF470	388566	NP_001001668	Y713	ILSSALPYHQVL---	TNK1
ZNF670	93474	NP_149990	Y281	THTGEKPYECIKCGK	FynTR
ZNF706	51123	NP_057180	Y39	AAKAALIYTCTVCRT	FES
ZNF787	126208	NP_001002836	Y122	IHTGEKPYACLECGK	FynTR
ZNRF3	84133	NP_115549	Y499	SDYDPFIYRSRSPCR	SRC
ZRANB2	9406	NP_976225	Y124	REESDGEYDEFGRKK	FGR

7. References

AHMAD, S. S., GLATZLE, J., BAJAEIFER, K., BUHLER, S., LEHMANN, T., KONIGSRAINER, I., VOLLMER, J. P., SIPOS, B., AHMAD, S. S., NORTHOFF, H., KONIGSRAINER, A. & ZIEKER, D. 2013. Phosphoglycerate kinase 1 as a promoter of metastasis in colon cancer. *Int J Oncol,* 43**,** 586-90.

ALBERT, R. 2005. Scale-free networks in cell biology. *J Cell Sci,* 118**,** 4947-57.

ALEXANDER, J., LIM, D., JOUGHIN, B. A., HEGEMANN, B., HUTCHINS, J. R., EHRENBERGER, T., IVINS, F., SESSA, F., HUDECZ, O., NIGG, E. A., FRY, A. M., MUSACCHIO, A., STUKENBERG, P. T., MECHTLER, K., PETERS, J. M., SMERDON, S. J. & YAFFE, M. B. 2011. Spatial exclusivity combined with positive and negative selection of phosphorylation motifs is the basis for context-dependent mitotic signaling. *Sci Signal,* 4**,** ra42.

ALONSO, A., SASIN, J., BOTTINI, N., FRIEDBERG, I., FRIEDBERG, I., OSTERMAN, A., GODZIK, A., HUNTER, T., DIXON, J. & MUSTELIN, T. 2004. Protein tyrosine phosphatases in the human genome. *Cell,* 117**,** 699-711.

AMANCHY, R., PERIASWAMY, B., MATHIVANAN, S., REDDY, R., TATTIKOTA, S. G. & PANDEY, A. 2007. A curated compendium of phosphorylation motifs. *Nat Biotechnol,* 25**,** 285-6.

APELT, F. 2012. Revealing Protein Tyrosine Kinase Phosphorylation Specificity in Yeast Interactome Networks. *In:* POTSDAM, U. O. (ed.).

ARNAUD, L., BALLIF, B. A., FORSTER, E. & COOPER, J. A. 2003. Fyn tyrosine kinase is a critical regulator of disabled-1 during brain development. *Curr Biol,* 13**,** 9-17.

BALDI, P., BRUNAK, S., CHAUVIN, Y., ANDERSEN, C. A. & NIELSEN, H. 2000. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics,* 16**,** 412-24.

BALLIF, B. A., ARNAUD, L., ARTHUR, W. T., GURIS, D., IMAMOTO, A. & COOPER, J. A. 2004. Activation of a Dab1/CrkL/C3G/Rap1 pathway in Reelin-stimulated neurons. *Curr Biol,* 14**,** 606-10.

BALLIF, B. A., CAREY, G. R., SUNYAEV, S. R. & GYGI, S. P. 2008. Large-scale identification and evolution indexing of tyrosine phosphorylation sites from murine brain. *J Proteome Res,* 7**,** 311-8.

BANDURA, D. R., BARANOV, V. I., ORNATSKY, O. I., ANTONOV, A., KINACH, R., LOU, X., PAVLOV, S., VOROBIEV, S., DICK, J. E. & TANNER, S. D. 2009. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem,* 81**,** 6813-22.

BANTSCHEFF, M., EBERHARD, D., ABRAHAM, Y., BASTUCK, S., BOESCHE, M., HOBSON, S., MATHIESON, T., PERRIN, J., RAIDA, M., RAU, C., READER, V., SWEETMAN, G., BAUER, A., BOUWMEESTER, T., HOPF, C., KRUSE, U., NEUBAUER, G., RAMSDEN, N., RICK, J., KUSTER, B. & DREWES, G. 2007. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol,* 25**,** 1035-44.

BEAUSOLEIL, S. A., VILLEN, J., GERBER, S. A., RUSH, J. & GYGI, S. P. 2006. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol,* 24**,** 1285-92.

BECKER, W. & JOOST, H. G. 1999. Structural and functional characteristics of Dyrk, a novel subfamily of protein kinases with dual specificity. *Prog Nucleic Acid Res Mol Biol,* 62**,** 1-17.

BEWARDER, N., WEINRICH, V., BUDDE, P., HARTMANN, D., FLASWINKEL, H., RETH, M. & FREY, J. 1996. In vivo and in vitro specificity of protein tyrosine kinases for immunoglobulin G receptor (FcgammaRII) phosphorylation. *Mol Cell Biol,* 16**,** 4735-43.

BLOM, N., GAMMELTOFT, S. & BRUNAK, S. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol,* 294**,** 1351-62.

BLOM, N., SICHERITZ-PONTEN, T., GUPTA, R., GAMMELTOFT, S. & BRUNAK, S. 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics,* 4**,** 1633-49.

BLUEMLEIN, K., GRUNING, N. M., FEICHTINGER, R. G., LEHRACH, H., KOFLER, B. & RALSER, M. 2011. No evidence for a shift in pyruvate kinase PKM1 to PKM2 expression during tumorigenesis. *Oncotarget,* 2**,** 393-400.

BLUME-JENSEN, P. & HUNTER, T. 2001. Oncogenic kinase signalling. *Nature,* 411**,** 355-65.
BODENMILLER, B., WANKA, S., KRAFT, C., URBAN, J., CAMPBELL, D., PEDRIOLI, P. G., GERRITS, B., PICOTTI, P., LAM, H., VITEK, O., BRUSNIAK, M. Y., ROSCHITZKI, B., ZHANG, C., SHOKAT, K. M., SCHLAPBACH, R., COLMAN-LERNER, A., NOLAN, G. P., NESVIZHSKII, A. I., PETER, M., LOEWITH, R., VON MERING, C. & AEBERSOLD, R. 2010. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Sci Signal,* 3**,** rs4.
BOERSEMA, P. J., FOONG, L. Y., DING, V. M., LEMEER, S., VAN BREUKELEN, B., PHILP, R., BOEKHORST, J., SNEL, B., DEN HERTOG, J., CHOO, A. B. & HECK, A. J. 2010. In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. *Mol Cell Proteomics,* 9**,** 84-99.
BOGGON, T. J. & ECK, M. J. 2004. Structure and regulation of Src family kinases. *Oncogene,* 23**,** 7918-27.
BOSCHELLI, F., UPTAIN, S. M. & LIGHTBODY, J. J. 1993. The lethality of p60v-src in Saccharomyces cerevisiae and the activation of p34CDC28 kinase are dependent on the integrity of the SH2 domain. *J Cell Sci,* 105 (Pt 2)**,** 519-28.
BOTSTEIN, D., CHERVITZ, S. A. & CHERRY, J. M. 1997. Yeast as a model organism. *Science,* 277**,** 1259-60.
BOYLE, E. I., WENG, S., GOLLUB, J., JIN, H., BOTSTEIN, D., CHERRY, J. M. & SHERLOCK, G. 2004. GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics,* 20**,** 3710-5.
BRADSHAW, J. M. 2010. The Src, Syk, and Tec family kinases: distinct types of molecular switches. *Cell Signal,* 22**,** 1175-84.
BREITKREUTZ, A., CHOI, H., SHAROM, J. R., BOUCHER, L., NEDUVA, V., LARSEN, B., LIN, Z. Y., BREITKREUTZ, B. J., STARK, C., LIU, G., AHN, J., DEWAR-DARCH, D., REGULY, T., TANG, X., ALMEIDA, R., QIN, Z. S., PAWSON, T., GINGRAS, A. C., NESVIZHSKII, A. I. & TYERS, M. 2010. A global protein kinase and phosphatase interaction network in yeast. *Science,* 328**,** 1043-6.
BROWN, M. T. & COOPER, J. A. 1996. Regulation, substrates and functions of src. *Biochim Biophys Acta,* 1287**,** 121-49.
BRUGGE, J. S., JAROSIK, G., ANDERSEN, J., QUERAL-LUSTIG, A., FEDOR-CHAIKEN, M. & BROACH, J. R. 1987. Expression of Rous sarcoma virus transforming protein pp60v-src in Saccharomyces cerevisiae cells. *Mol Cell Biol,* 7**,** 2180-7.
BUSZEWSKI, B. & NOGA, S. 2012. Hydrophilic interaction liquid chromatography (HILIC)--a powerful separation technique. *Anal Bioanal Chem,* 402**,** 231-47.
CALALB, M. B., POLTE, T. R. & HANKS, S. K. 1995. Tyrosine phosphorylation of focal adhesion kinase at sites in the catalytic domain regulates kinase activity: a role for Src family kinases. *Mol Cell Biol,* 15**,** 954-63.
CAMACHO-CARVAJAL, M. M., WOLLSCHEID, B., AEBERSOLD, R., STEIMLE, V. & SCHAMEL, W. W. 2004. Two-dimensional Blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates: a proteomics approach. *Mol Cell Proteomics,* 3**,** 176-82.
CAMPS, M., NICHOLS, A. & ARKINSTALL, S. 2000. Dual specificity phosphatases: a gene family for control of MAP kinase function. *FASEB J,* 14**,** 6-16.
CASCANTE, M., CENTELLES, J. J., VEECH, R. L., LEE, W. N. & BOROS, L. G. 2000. Role of thiamin (vitamin B-1) and transketolase in tumor cell proliferation. *Nutr Cancer,* 36**,** 150-4.
CHEERATHODI, M., VINCENT, J. J. & BALLIF, B. A. 2015. Quantitative comparison of CrkL-SH3 binding proteins from embryonic murine brain and liver: Implications for developmental signaling and the quantification of protein species variants in bottom-up proteomics. *J Proteomics,* 125**,** 104-11.

CHENG, H., ROGERS, J. A., DUNHAM, N. A. & SMITHGALL, T. E. 1999. Regulation of c-Fes tyrosine kinase and biological activities by N-terminal coiled-coil oligomerization domains. *Mol Cell Biol,* 19**,** 8335-43.

CHERRY, J. M., HONG, E. L., AMUNDSEN, C., BALAKRISHNAN, R., BINKLEY, G., CHAN, E. T., CHRISTIE, K. R., COSTANZO, M. C., DWIGHT, S. S., ENGEL, S. R., FISK, D. G., HIRSCHMAN, J. E., HITZ, B. C., KARRA, K., KRIEGER, C. J., MIYASATO, S. R., NASH, R. S., PARK, J., SKRZYPEK, M. S., SIMISON, M., WENG, S. & WONG, E. D. 2012. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res,* 40**,** D700-5.

CHOU, M. F., PRISIC, S., LUBNER, J. M., CHURCH, G. M., HUSSON, R. N. & SCHWARTZ, D. 2012. Using bacteria to determine protein kinase specificity and predict target substrates. *PLoS One,* 7**,** e52747.

CHOU, M. F. & SCHWARTZ, D. 2011. Biological sequence motif discovery using motif-x. *Curr Protoc Bioinformatics,* Chapter 13**,** Unit 13 15-24.

CHOUDHARY, C. & MANN, M. 2010. Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol,* 11**,** 427-39.

CHRISTOFK, H. R., VANDER HEIDEN, M. G., HARRIS, M. H., RAMANATHAN, A., GERSZTEN, R. E., WEI, R., FLEMING, M. D., SCHREIBER, S. L. & CANTLEY, L. C. 2008a. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature,* 452**,** 230-3.

CHRISTOFK, H. R., VANDER HEIDEN, M. G., WU, N., ASARA, J. M. & CANTLEY, L. C. 2008b. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature,* 452**,** 181-6.

CIESLA, J., FRACZYK, T. & RODE, W. 2011. Phosphorylation of basic amino acid residues in proteins: important but easily missed. *Acta Biochim Pol,* 58**,** 137-48.

COCHRANE, D., WEBSTER, C., MASIH, G. & MCCAFFERTY, J. 2000. Identification of natural ligands for SH2 domains from a phage display cDNA library. *J Mol Biol,* 297**,** 89-97.

COLAERT, N., HELSENS, K., MARTENS, L., VANDEKERCKHOVE, J. & GEVAERT, K. 2009. Improved visualization of protein consensus sequences by iceLogo. *Nat Methods,* 6**,** 786-7.

COLICELLI, J. 2010. ABL tyrosine kinases: evolution of function, regulation, and specificity. *Sci Signal,* 3**,** re6.

COMB, M. J. M., MA, US), TAN, YI (LYNNFIELD, MA, US). 2008. *Production of motif-specific and context-independent antibodies using peptide libraries as antigens*. United States patent application 7344714.

COOPER, J. A., ESCH, F. S., TAYLOR, S. S. & HUNTER, T. 1984. Phosphorylation sites in enolase and lactate dehydrogenase utilized by tyrosine protein kinases in vivo and in vitro. *J Biol Chem,* 259**,** 7835-41.

CORSON, L. B., YAMANAKA, Y., LAI, K. M. & ROSSANT, J. 2003. Spatial and temporal patterns of ERK signaling during mouse embryogenesis. *Development,* 130**,** 4527-37.

COX, J. & MANN, M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol,* 26**,** 1367-72.

COX, J., NEUHAUSER, N., MICHALSKI, A., SCHELTEMA, R. A., OLSEN, J. V. & MANN, M. 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res,* 10**,** 1794-805.

CRAIG, R. & BEAVIS, R. C. 2004. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics,* 20**,** 1466-7.

CROOKS, G. E., HON, G., CHANDONIA, J. M. & BRENNER, S. E. 2004. WebLogo: a sequence logo generator. *Genome Res,* 14**,** 1188-90.

CUJEC, T. P., MEDEIROS, P. F., HAMMOND, P., RISE, C. & KREIDER, B. L. 2002. Selection of v-abl tyrosine kinase substrate sequences from randomized peptide and cellular proteomic libraries using mRNA display. *Chem Biol,* 9**,** 253-64.

DAMLE, N. P. & MOHANTY, D. 2014. Deciphering kinase-substrate relationships by analysis of domain-specific phosphorylation network. *Bioinformatics,* 30**,** 1730-8.

DE CASTRO, R. O., ZHANG, J., JAMUR, M. C., OLIVER, C. & SIRAGANIAN, R. P. 2010. Tyrosines in the carboxyl terminus regulate Syk kinase activity and function. *J Biol Chem,* 285**,** 26674-84.

DE GODOY, L. M., OLSEN, J. V., COX, J., NIELSEN, M. L., HUBNER, N. C., FROHLICH, F., WALTHER, T. C. & MANN, M. 2008. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature,* 455**,** 1251-4.

DEDECKER, P., MO, G. C., DERTINGER, T. & ZHANG, J. 2012. Widely accessible method for superresolution fluorescence imaging of living systems. *Proc Natl Acad Sci U S A,* 109**,** 10909-14.

DENG, Y., ALICEA-VELAZQUEZ, N. L., BANNWARTH, L., LEHTONEN, S. I., BOGGON, T. J., CHENG, H. C., HYTONEN, V. P. & TURK, B. E. 2014. Global analysis of human nonreceptor tyrosine kinase specificity using high-density Peptide microarrays. *J Proteome Res,* 13**,** 4339-46.

DEPHOURE, N., HOWSON, R. W., BLETHROW, J. D., SHOKAT, K. M. & O'SHEA, E. K. 2005. Combining chemical genetics and proteomics to identify protein kinase substrates. *Proc Natl Acad Sci U S A,* 102**,** 17940-5.

DESIERE, F., DEUTSCH, E. W., KING, N. L., NESVIZHSKII, A. I., MALLICK, P., ENG, J., CHEN, S., EDDES, J., LOEVENICH, S. N. & AEBERSOLD, R. 2006. The PeptideAtlas project. *Nucleic Acids Res,* 34**,** D655-8.

DHANASEKARAN, N. & PREMKUMAR REDDY, E. 1998. Signaling by dual specificity kinases. *Oncogene,* 17**,** 1447-55.

DINKEL, H., CHICA, C., VIA, A., GOULD, C. M., JENSEN, L. J., GIBSON, T. J. & DIELLA, F. 2011. Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic Acids Res,* 39**,** D261-7.

DOMON, B. & AEBERSOLD, R. 2006. Mass spectrometry and protein analysis. *Science,* 312**,** 212-7.

DOSZTANYI, Z., CSIZMOK, V., TOMPA, P. & SIMON, I. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics,* 21**,** 3433-4.

DUARTE, M. L., PENA, D. A., NUNES FERRAZ, F. A., BERTI, D. A., PASCHOAL SOBREIRA, T. J., COSTA-JUNIOR, H. M., ABDEL BAQUI, M. M., DISATNIK, M. H., XAVIER-NETO, J., LOPES DE OLIVEIRA, P. S. & SCHECHTMAN, D. 2014. Protein folding creates structure-based, noncontiguous consensus phosphorylation motifs recognized by kinases. *Sci Signal,* 7**,** ra105.

DUKE-COHAN, J. S., KANG, H., LIU, H. & RUDD, C. E. 2006. Regulation and function of SKAP-55 non-canonical motif binding to the SH3c domain of adhesion and degranulation-promoting adaptor protein. *J Biol Chem,* 281**,** 13743-50.

EJSING, C. S., SAMPAIO, J. L., SURENDRANATH, V., DUCHOSLAV, E., EKROOS, K., KLEMM, R. W., SIMONS, K. & SHEVCHENKO, A. 2009. Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proc Natl Acad Sci U S A,* 106**,** 2136-41.

ENG, J. K., MCCORMACK, A. L. & YATES, J. R. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom,* 5**,** 976-89.

ERPEL, T., SUPERTI-FURGA, G. & COURTNEIDGE, S. A. 1995. Mutational analysis of the Src SH3 domain: the same residues of the ligand binding surface are important for intra- and intermolecular interactions. *EMBO J,* 14**,** 963-75.

FAGERBERG, L., HALLSTROM, B. M., OKSVOLD, P., KAMPF, C., DJUREINOVIC, D., ODEBERG, J., HABUKA, M., TAHMASEBPOOR, S., DANIELSSON, A., EDLUND, K., ASPLUND, A., SJOSTEDT, E., LUNDBERG, E., SZIGYARTO, C. A., SKOGS, M., TAKANEN, J. O., BERLING, H., TEGEL, H., MULDER, J., NILSSON, P., SCHWENK, J. M., LINDSKOG, C., DANIELSSON, F., MARDINOGLU, A., SIVERTSSON, A., VON FEILITZEN, K., FORSBERG, M., ZWAHLEN, M., OLSSON, I., NAVANI, S., HUSS, M., NIELSEN, J., PONTEN, F. & UHLEN, M. 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics,* 13**,** 397-406.

FAN, J., KANG, H. B., SHAN, C., ELF, S., LIN, R., XIE, J., GU, T. L., AGUIAR, M., LONNING, S., CHUNG, T. W., ARELLANO, M., KHOURY, H. J., SHIN, D. M., KHURI, F. R., BOGGON, T. J., KANG, S. &

CHEN, J. 2014. Tyr-301 phosphorylation inhibits pyruvate dehydrogenase by blocking substrate binding and promotes the Warburg effect. *J Biol Chem,* 289**,** 26533-41.
FARRIOL-MATHIS, N., GARAVELLI, J. S., BOECKMANN, B., DUVAUD, S., GASTEIGER, E., GATEAU, A., VEUTHEY, A. L. & BAIROCH, A. 2004. Annotation of post-translational modifications in the Swiss-Prot knowledge base. *Proteomics,* 4**,** 1537-50.
FELLER, S. M., KNUDSEN, B. & HANAFUSA, H. 1994. c-Abl kinase regulates the protein binding activity of c-Crk. *EMBO J,* 13**,** 2341-51.
FICARRO, S., CHERTIHIN, O., WESTBROOK, V. A., WHITE, F., JAYES, F., KALAB, P., MARTO, J. A., SHABANOWITZ, J., HERR, J. C., HUNT, D. F. & VISCONTI, P. E. 2003. Phosphoproteome analysis of capacitated human sperm. Evidence of tyrosine phosphorylation of a kinase-anchoring protein 3 and valosin-containing protein/p97 during capacitation. *J Biol Chem,* 278**,** 11579-89.
FICARRO, S. B., MCCLELAND, M. L., STUKENBERG, P. T., BURKE, D. J., ROSS, M. M., SHABANOWITZ, J., HUNT, D. F. & WHITE, F. M. 2002. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol,* 20**,** 301-5.
FLORIO, M., WILSON, L. K., TRAGER, J. B., THORNER, J. & MARTIN, G. S. 1994. Aberrant protein phosphorylation at tyrosine is responsible for the growth-inhibitory action of pp60v-src expressed in the yeast Saccharomyces cerevisiae. *Mol Biol Cell,* 5**,** 283-96.
FRIEDMAN, A. A., TUCKER, G., SINGH, R., YAN, D., VINAYAGAM, A., HU, Y., BINARI, R., HONG, P., SUN, X., PORTO, M., PACIFICO, S., MURALI, T., FINLEY, R. L., JR., ASARA, J. M., BERGER, B. & PERRIMON, N. 2011. Proteomic and functional genomic landscape of receptor tyrosine kinase and ras to extracellular signal-regulated kinase signaling. *Sci Signal,* 4**,** rs10.
GAJIWALA, K. S., MAEGLEY, K., FERRE, R., HE, Y. A. & YU, X. 2013. Ack1: activation and regulation by allostery. *PLoS One,* 8**,** e53994.
GAO, Z. & GODBOUT, R. 2013. Reelin-Disabled-1 signaling in neuronal migration: splicing takes the stage. *Cell Mol Life Sci,* 70**,** 2319-29.
GAVIN, A. C., ALOY, P., GRANDI, P., KRAUSE, R., BOESCHE, M., MARZIOCH, M., RAU, C., JENSEN, L. J., BASTUCK, S., DUMPELFELD, B., EDELMANN, A., HEURTIER, M. A., HOFFMAN, V., HOEFERT, C., KLEIN, K., HUDAK, M., MICHON, A. M., SCHELDER, M., SCHIRLE, M., REMOR, M., RUDI, T., HOOPER, S., BAUER, A., BOUWMEESTER, T., CASARI, G., DREWES, G., NEUBAUER, G., RICK, J. M., KUSTER, B., BORK, P., RUSSELL, R. B. & SUPERTI-FURGA, G. 2006. Proteome survey reveals modularity of the yeast cell machinery. *Nature,* 440**,** 631-6.
GHAEMMAGHAMI, S., HUH, W. K., BOWER, K., HOWSON, R. W., BELLE, A., DEPHOURE, N., O'SHEA, E. K. & WEISSMAN, J. S. 2003. Global analysis of protein expression in yeast. *Nature,* 425**,** 737-41.
GLENNEY, J. R., JR., ZOKAS, L. & KAMPS, M. P. 1988. Monoclonal antibodies to phosphotyrosine. *J Immunol Methods,* 109**,** 277-85.
GNAD, F., DE GODOY, L. M., COX, J., NEUHAUSER, N., REN, S., OLSEN, J. V. & MANN, M. 2009. High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. *Proteomics,* 9**,** 4642-52.
GOH, K. I., OH, E., JEONG, H., KAHNG, B. & KIM, D. 2002. Classification of scale-free networks. *Proc Natl Acad Sci U S A,* 99**,** 12583-8.
GOLDBERG, J. M., MANNING, G., LIU, A., FEY, P., PILCHER, K. E., XU, Y. & SMITH, J. L. 2006. The dictyostelium kinome--analysis of the protein kinases from a simple model organism. *PLoS Genet,* 2**,** e38.
GONFLONI, S., WILLIAMS, J. C., HATTULA, K., WEIJLAND, A., WIERENGA, R. K. & SUPERTI-FURGA, G. 1997. The role of the linker between the SH2 domain and catalytic domain in the regulation and function of Src. *EMBO J,* 16**,** 7261-71.
GRABLY, M. & ENGELBERG, D. 2010. A detailed protocol for chromatin immunoprecipitation in the yeast Saccharomyces cerevisiae. *Methods Mol Biol,* 638**,** 211-24.

GRADLER, U., SCHWARZ, D., DRESING, V., MUSIL, D., BOMKE, J., FRECH, M., GREINER, H., JAKEL, S., RYSIOK, T., MULLER-POMPALLA, D. & WEGENER, A. 2013. Structural and biophysical characterization of the Syk activation switch. *J Mol Biol,* 425**,** 309-33.

GREER, P. 2002. Closing in on the biological functions of Fps/Fes and Fer. *Nat Rev Mol Cell Biol,* 3**,** 278-89.

GRIMMLER, M., WANG, Y., MUND, T., CILENSEK, Z., KEIDEL, E. M., WADDELL, M. B., JAKEL, H., KULLMANN, M., KRIWACKI, R. W. & HENGST, L. 2007. Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases. *Cell,* 128**,** 269-80.

GROSSMANN, A., BENLASFER, N., BIRTH, P., HEGELE, A., WACHSMUTH, F., APELT, L. & STELZL, U. 2015. Phospho-tyrosine dependent protein-protein interaction network. *Mol Syst Biol,* 11**,** 794.

GUNDE, T. & BARBERIS, A. 2005. Yeast growth selection system for detecting activity and inhibition of dimerization-dependent receptor tyrosine kinase. *Biotechniques,* 39**,** 541-9.

HAGBERG, A. A., SCHULT, D. A. & SWART, P. J. 2008. Exploring Network Structure, Dynamics, and Function using NetworkX. *Proceedings of the 7th Python in Science Conference.*

HARRIS, L. K., FRUMM, S. M. & BISHOP, A. C. 2013. A general assay for monitoring the activities of protein tyrosine phosphatases in living eukaryotic cells. *Anal Biochem,* 435**,** 99-105.

HEISTERKAMP, N., JENSTER, G., TEN HOEVE, J., ZOVICH, D., PATTENGALE, P. K. & GROFFEN, J. 1990. Acute leukaemia in bcr/abl transgenic mice. *Nature,* 344**,** 251-3.

HINKLE, K., WEIR, M., FULTON, Z., HAO, J., MANN, J., MCGEHEE, A., CORWIN, T., STELZL, U., DEMING, P., JUO, P. & BALLIF, B. 2015. Novel Tyrosine Phosphorylation Sites Fine Tune the Activity and Substrate Binding of Src Family Kinases. *The FASEB Journal,* 29.

HITOSUGI, T. & CHEN, J. 2014. Post-translational modifications and the Warburg effect. *Oncogene,* 33**,** 4279-85.

HITOSUGI, T., KANG, S., VANDER HEIDEN, M. G., CHUNG, T. W., ELF, S., LYTHGOE, K., DONG, S., LONIAL, S., WANG, X., CHEN, G. Z., XIE, J., GU, T. L., POLAKIEWICZ, R. D., ROESEL, J. L., BOGGON, T. J., KHURI, F. R., GILLILAND, D. G., CANTLEY, L. C., KAUFMAN, J. & CHEN, J. 2009. Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth. *Sci Signal,* 2**,** ra73.

HITOSUGI, T., ZHOU, L., ELF, S., FAN, J., KANG, H. B., SEO, J. H., SHAN, C., DAI, Q., ZHANG, L., XIE, J., GU, T. L., JIN, P., ALECKOVIC, M., LEROY, G., KANG, Y., SUDDERTH, J. A., DEBERARDINIS, R. J., LUAN, C. H., CHEN, G. Z., MULLER, S., SHIN, D. M., OWONIKOKO, T. K., LONIAL, S., ARELLANO, M. L., KHOURY, H. J., KHURI, F. R., LEE, B. H., YE, K., BOGGON, T. J., KANG, S., HE, C. & CHEN, J. 2012. Phosphoglycerate mutase 1 coordinates glycolysis and biosynthesis to promote tumor growth. *Cancer Cell,* 22**,** 585-600.

HITOSUGI, T., ZHOU, L., FAN, J., ELF, S., ZHANG, L., XIE, J., WANG, Y., GU, T. L., ALECKOVIC, M., LEROY, G., KANG, Y., KANG, H. B., SEO, J. H., SHAN, C., JIN, P., GONG, W., LONIAL, S., ARELLANO, M. L., KHOURY, H. J., CHEN, G. Z., SHIN, D. M., KHURI, F. R., BOGGON, T. J., KANG, S., HE, C. & CHEN, J. 2013. Tyr26 phosphorylation of PGAM1 provides a metabolic advantage to tumours by stabilizing the active conformation. *Nat Commun,* 4**,** 1790.

HJERMSTAD, S. J., PETERS, K. L., BRIGGS, S. D., GLAZER, R. I. & SMITHGALL, T. E. 1993. Regulation of the human c-fes protein tyrosine kinase (p93c-fes) by its src homology 2 domain and major autophosphorylation site (Tyr-713). *Oncogene,* 8**,** 2283-92.

HOLT, L. J., TUCH, B. B., VILLEN, J., JOHNSON, A. D., GYGI, S. P. & MORGAN, D. O. 2009. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science,* 325**,** 1682-6.

HONG, E., SHIN, J., KIM, H. I., LEE, S. T. & LEE, W. 2004. Solution structure and backbone dynamics of the non-receptor protein-tyrosine kinase-6 Src homology 2 domain. *J Biol Chem,* 279**,** 29700-8.

HORN, H., SCHOOF, E. M., KIM, J., ROBIN, X., MILLER, M. L., DIELLA, F., PALMA, A., CESARENI, G., JENSEN, L. J. & LINDING, R. 2014. KinomeXplorer: an integrated platform for kinome biology studies. *Nat Methods,* 11**,** 603-4.

HORNBECK, P. V., KORNHAUSER, J. M., TKACHEV, S., ZHANG, B., SKRZYPEK, E., MURRAY, B., LATHAM, V. & SULLIVAN, M. 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**,** D261-70.

HU, H., BLISS, J. M., WANG, Y. & COLICELLI, J. 2005. RIN1 is an ABL tyrosine kinase activator and a regulator of epithelial-cell adhesion and migration. *Curr Biol,* 15**,** 815-23.

HU, J., LOCASALE, J. W., BIELAS, J. H., O'SULLIVAN, J., SHEAHAN, K., CANTLEY, L. C., VANDER HEIDEN, M. G. & VITKUP, D. 2013. Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat Biotechnol,* 31**,** 522-9.

HUA, L., ZHU, M., SONG, X., WANG, J., FANG, Z., ZHANG, C., SHI, Q., ZHAN, W., WANG, L., MENG, Q., ZHOU, X. & YU, R. 2014. FRK suppresses the proliferation of human glioma cells by inhibiting cyclin D1 nuclear accumulation. *J Neurooncol,* 119**,** 49-58.

HUANG, H. D., LEE, T. Y., TZENG, S. W. & HORNG, J. T. 2005. KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic Acids Res,* 33**,** W226-9.

HUERTA-CEPAS, J., DOPAZO, J. & GABALDON, T. 2010. ETE: a python Environment for Tree Exploration. *BMC Bioinformatics,* 11**,** 24.

HUH, W. K., FALVO, J. V., GERKE, L. C., CARROLL, A. S., HOWSON, R. W., WEISSMAN, J. S. & O'SHEA, E. K. 2003. Global analysis of protein localization in budding yeast. *Nature,* 425**,** 686-91.

HUNTER, T. 2012. Why nature chose phosphate to modify proteins. *Philos Trans R Soc Lond B Biol Sci,* 367**,** 2513-6.

HUNTER, T. & SEFTON, B. M. 1980. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci U S A,* 77**,** 1311-5.

HUTTI, J. E., JARRELL, E. T., CHANG, J. D., ABBOTT, D. W., STORZ, P., TOKER, A., CANTLEY, L. C. & TURK, B. E. 2004. A rapid method for determining protein kinase phosphorylation specificity. *Nat Methods,* 1**,** 27-9.

HUTTLIN, E. L., JEDRYCHOWSKI, M. P., ELIAS, J. E., GOSWAMI, T., RAD, R., BEAUSOLEIL, S. A., VILLEN, J., HAAS, W., SOWA, M. E. & GYGI, S. P. 2010. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell,* 143**,** 1174-89.

IAKOUCHEVA, L. M., RADIVOJAC, P., BROWN, C. J., O'CONNOR, T. R., SIKES, J. G., OBRADOVIC, Z. & DUNKER, A. K. 2004. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res,* 32**,** 1037-49.

JIN, J. & PAWSON, T. 2012. Modular evolution of phosphorylation-based signalling systems. *Philos Trans R Soc Lond B Biol Sci,* 367**,** 2540-55.

JOHNSON, L. N., NOBLE, M. E. & OWEN, D. J. 1996. Active and inactive protein kinases: structural basis for regulation. *Cell,* 85**,** 149-58.

JOSEPH, R. E. & ANDREOTTI, A. H. 2009. Conformational snapshots of Tec kinases during signaling. *Immunol Rev,* 228**,** 74-92.

JOUGHIN, B. A., LIU, C., LAUFFENBURGER, D. A., HOGUE, C. W. & YAFFE, M. B. 2012. Protein kinases display minimal interpositional dependence on substrate sequence: potential implications for the evolution of signalling networks. *Philos Trans R Soc Lond B Biol Sci,* 367**,** 2574-83.

JOY, M. P., BROCK, A., INGBER, D. E. & HUANG, S. 2005. High-betweenness proteins in the yeast protein interaction network. *J Biomed Biotechnol,* 2005**,** 96-103.

JUNGER, M. A. & AEBERSOLD, R. 2014. Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. *Wiley Interdiscip Rev Dev Biol,* 3**,** 83-112.

KACHROO, A. H., LAURENT, J. M., YELLMAN, C. M., MEYER, A. G., WILKE, C. O. & MARCOTTE, E. M. 2015. Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science,* 348**,** 921-5.

KAMBUROV, A., STELZL, U., LEHRACH, H. & HERWIG, R. 2013. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res,* 41**,** D793-800.

KANAKURA, Y., DRUKER, B., CANNISTRA, S. A., FURUKAWA, Y., TORIMOTO, Y. & GRIFFIN, J. D. 1990. Signal transduction of the human granulocyte-macrophage colony-stimulating factor and

interleukin-3 receptors involves tyrosine phosphorylation of a common set of cytoplasmic proteins. *Blood,* 76**,** 706-15.

KIM, J. H., LEE, J., OH, B., KIMM, K. & KOH, I. 2004. Prediction of phosphorylation sites using SVMs. *Bioinformatics,* 20**,** 3179-84.

KING, N. L., DEUTSCH, E. W., RANISH, J. A., NESVIZHSKII, A. I., EDDES, J. S., MALLICK, P., ENG, J., DESIERE, F., FLORY, M., MARTIN, D. B., KIM, B., LEE, H., RAUGHT, B. & AEBERSOLD, R. 2006. Analysis of the Saccharomyces cerevisiae proteome with PeptideAtlas. *Genome Biol,* 7**,** R106.

KO, S., AHN, K. E., LEE, Y. M., AHN, H. C. & LEE, W. 2009. Structural basis of the auto-inhibition mechanism of nonreceptor tyrosine kinase PTK6. *Biochem Biophys Res Commun,* 384**,** 236-42.

KOBAYASHI, H., SAITO, T., SATO, K., FURUSAWA, K., HOSOKAWA, T., TSUTSUMI, K., ASADA, A., KAMADA, S., OHSHIMA, T. & HISANAGA, S. 2014. Phosphorylation of cyclin-dependent kinase 5 (Cdk5) at Tyr-15 is inhibited by Cdk5 activators and does not contribute to the activation of Cdk5. *J Biol Chem,* 289**,** 19627-36.

KOEGL, M., COURTNEIDGE, S. A. & SUPERTI-FURGA, G. 1995. Structural requirements for the efficient regulation of the Src protein tyrosine kinase by Csk. *Oncogene,* 11**,** 2317-29.

KORNBLUTH, S., JOVE, R. & HANAFUSA, H. 1987. Characterization of avian and viral p60src proteins expressed in yeast. *Proc Natl Acad Sci U S A,* 84**,** 4455-9.

KOYAMA, M., SAITO, S., NAKAGAWA, R., KATSUYAMA, I., HATANAKA, M., YAMAMOTO, T., ARAKAWA, T. & TOKUNAG, M. 2006. Expression of human tyrosine kinase, Lck, in yeast Saccharomyces cerevisiae: growth suppression and strategy for inhibitor screening. *Protein Pept Lett,* 13**,** 915-20.

KREEGIPUU, A., BLOM, N., BRUNAK, S. & JARV, J. 1998. Statistical analysis of protein kinase specificity determinants. *FEBS Lett,* 430**,** 45-50.

KROGAN, N. J., CAGNEY, G., YU, H., ZHONG, G., GUO, X., IGNATCHENKO, A., LI, J., PU, S., DATTA, N., TIKUISIS, A. P., PUNNA, T., PEREGRIN-ALVAREZ, J. M., SHALES, M., ZHANG, X., DAVEY, M., ROBINSON, M. D., PACCANARO, A., BRAY, J. E., SHEUNG, A., BEATTIE, B., RICHARDS, D. P., CANADIEN, V., LALEV, A., MENA, F., WONG, P., STAROSTINE, A., CANETE, M. M., VLASBLOM, J., WU, S., ORSI, C., COLLINS, S. R., CHANDRAN, S., HAW, R., RILSTONE, J. J., GANDI, K., THOMPSON, N. J., MUSSO, G., ST ONGE, P., GHANNY, S., LAM, M. H., BUTLAND, G., ALTAF-UL, A. M., KANAYA, S., SHILATIFARD, A., O'SHEA, E., WEISSMAN, J. S., INGLES, C. J., HUGHES, T. R., PARKINSON, J., GERSTEIN, M., WODAK, S. J., EMILI, A. & GREENBLATT, J. F. 2006. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature,* 440**,** 637-43.

KRUGER, M., KRATCHMAROVA, I., BLAGOEV, B., TSENG, Y. H., KAHN, C. R. & MANN, M. 2008. Dissection of the insulin signaling pathway via quantitative phosphoproteomics. *Proc Natl Acad Sci U S A,* 105**,** 2451-6.

KUMAR, S., FAJARDO, J. E., BIRGE, R. B. & SRIRAM, G. 2014. Crk at the quarter century mark: perspectives in signaling and cancer. *J Cell Biochem,* 115**,** 819-25.

LARSEN, M. R., THINGHOLM, T. E., JENSEN, O. N., ROEPSTORFF, P. & JORGENSEN, T. J. 2005. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics,* 4**,** 873-86.

LERNER, E. C., TRIBLE, R. P., SCHIAVONE, A. P., HOCHREIN, J. M., ENGEN, J. R. & SMITHGALL, T. E. 2005. Activation of the Src family kinase Hck without SH3-linker release. *J Biol Chem,* 280**,** 40832-7.

LIETHA, D., CAI, X., CECCARELLI, D. F., LI, Y., SCHALLER, M. D. & ECK, M. J. 2007. Structural basis for the autoinhibition of focal adhesion kinase. *Cell,* 129**,** 1177-87.

LIN, Y. H., PARK, Z. Y., LIN, D., BRAHMBHATT, A. A., RIO, M. C., YATES, J. R., 3RD & KLEMKE, R. L. 2004. Regulation of cell migration and survival by focal adhesion targeting of Lasp-1. *J Cell Biol,* 165**,** 421-32.

LINDING, R., JENSEN, L. J., DIELLA, F., BORK, P., GIBSON, T. J. & RUSSELL, R. B. 2003. Protein disorder prediction: implications for structural proteomics. *Structure,* 11**,** 1453-9.

LINDING, R., JENSEN, L. J., OSTHEIMER, G. J., VAN VUGT, M. A., JORGENSEN, C., MIRON, I. M., DIELLA, F., COLWILL, K., TAYLOR, L., ELDER, K., METALNIKOV, P., NGUYEN, V., PASCULESCU, A., JIN, J., PARK, J. G., SAMSON, L. D., WOODGETT, J. R., RUSSELL, R. B., BORK, P., YAFFE, M. B. & PAWSON, T. 2007. Systematic discovery of in vivo phosphorylation networks. *Cell,* 129**,** 1415-26.

LIU, B. A., JABLONOWSKI, K., SHAH, E. E., ENGELMANN, B. W., JONES, R. B. & NASH, P. D. 2010. SH2 domains recognize contextual peptide sequence information to determine selectivity. *Mol Cell Proteomics,* 9**,** 2391-404.

LIU, X., YU, X., ZACK, D. J., ZHU, H. & QIAN, J. 2008. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics,* 9**,** 271.

LOCK, P., ABRAM, C. L., GIBSON, T. & COURTNEIDGE, S. A. 1998. A new method for isolating tyrosine kinase substrates used to identify fish, an SH3 and PX domain-containing protein, and Src substrate. *EMBO J,* 17**,** 4346-57.

MANDINE, E., JEAN-BAPTISTE, V., VAYSSIERE, B., GOFFLO, D., BENARD, D., SARUBBI, E., DEPREZ, P., BARON, R., SUPERTI-FURGA, G. & LESUISSE, D. 2002. High-affinity Src-SH2 ligands which do not activate Tyr(527)-phosphorylated Src in an experimental in vivo system. *Biochem Biophys Res Commun,* 298**,** 185-92.

MANNING, B. D. & CANTLEY, L. C. 2002. Hitting the target: emerging technologies in the search for kinase substrates. *Sci STKE,* 2002**,** pe49.

MANNING, G., PLOWMAN, G. D., HUNTER, T. & SUDARSANAM, S. 2002a. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci,* 27**,** 514-20.

MANNING, G., WHYTE, D. B., MARTINEZ, R., HUNTER, T. & SUDARSANAM, S. 2002b. The protein kinase complement of the human genome. *Science,* 298**,** 1912-34.

MAROUGA, R., DAVID, S. & HAWKINS, E. 2005. The development of the DIGE system: 2D fluorescence difference gel analysis technology. *Anal Bioanal Chem,* 382**,** 669-78.

MATSUOKA, S., BALLIF, B. A., SMOGORZEWSKA, A., MCDONALD, E. R., 3RD, HUROV, K. E., LUO, J., BAKALARSKI, C. E., ZHAO, Z., SOLIMINI, N., LERENTHAL, Y., SHILOH, Y., GYGI, S. P. & ELLEDGE, S. J. 2007. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science,* 316**,** 1160-6.

MAZUREK, S. 2011. Pyruvate kinase type M2: a key regulator of the metabolic budget system in tumor cells. *Int J Biochem Cell Biol,* 43**,** 969-80.

MAZUREK, S., BOSCHEK, C. B., HUGO, F. & EIGENBRODT, E. 2005. Pyruvate kinase type M2 and its role in tumor growth and spreading. *Semin Cancer Biol,* 15**,** 300-8.

MI, T., MERLIN, J. C., DEVERASETTY, S., GRYK, M. R., BILL, T. J., BROOKS, A. W., LEE, L. Y., RATHNAYAKE, V., ROSS, C. A., SARGEANT, D. P., STRONG, C. L., WATTS, P., RAJASEKARAN, S. & SCHILLER, M. R. 2012. Minimotif Miner 3.0: database expansion and significantly improved reduction of false-positive predictions from consensus sequences. *Nucleic Acids Res,* 40**,** D252-60.

MILLER, M. L., JENSEN, L. J., DIELLA, F., JORGENSEN, C., TINTI, M., LI, L., HSIUNG, M., PARKER, S. A., BORDEAUX, J., SICHERITZ-PONTEN, T., OLHOVSKY, M., PASCULESCU, A., ALEXANDER, J., KNAPP, S., BLOM, N., BORK, P., LI, S., CESARENI, G., PAWSON, T., TURK, B. E., YAFFE, M. B., BRUNAK, S. & LINDING, R. 2008. Linear motif atlas for phosphorylation-dependent signaling. *Sci Signal,* 1**,** ra2.

MOK, J., KIM, P. M., LAM, H. Y., PICCIRILLO, S., ZHOU, X., JESCHKE, G. R., SHERIDAN, D. L., PARKER, S. A., DESAI, V., JWA, M., CAMERONI, E., NIU, H., GOOD, M., REMENYI, A., MA, J. L., SHEU, Y. J., SASSI, H. E., SOPKO, R., CHAN, C. S., DE VIRGILIO, C., HOLLINGSWORTH, N. M., LIM, W. A., STERN, D. F., STILLMAN, B., ANDREWS, B. J., GERSTEIN, M. B., SNYDER, M. & TURK, B. E. 2010. Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. *Sci Signal,* 3**,** ra12.

MONTALIBET, J. & KENNEDY, B. P. 2004. Using yeast to screen for inhibitors of protein tyrosine phosphatase 1B. *Biochem Pharmacol,* 68**,** 1807-14.

MOORHEAD, G. B., DE WEVER, V., TEMPLETON, G. & KERK, D. 2009. Evolution of protein phosphatases in plants and animals. *Biochem J,* 417**,** 401-9.
MORANDELL, S., GROSSTESSNER-HAIN, K., ROITINGER, E., HUDECZ, O., LINDHORST, T., TEIS, D., WRULICH, O. A., MAZANEK, M., TAUS, T., UEBERALL, F., MECHTLER, K. & HUBER, L. A. 2010. QIKS--Quantitative identification of kinase substrates. *Proteomics,* 10**,** 2015-25.
MURPHY, S. M., BERGMAN, M. & MORGAN, D. O. 1993. Suppression of c-Src activity by C-terminal Src kinase involves the c-Src SH2 and SH3 domains: analysis with Saccharomyces cerevisiae. *Mol Cell Biol,* 13**,** 5290-300.
NADA, S., OKADA, M., MACAULEY, A., COOPER, J. A. & NAKAGAWA, H. 1991. Cloning of a complementary DNA for a protein-tyrosine kinase that specifically phosphorylates a negative regulatory site of p60c-src. *Nature,* 351**,** 69-72.
NAGAR, B., HANTSCHEL, O., YOUNG, M. A., SCHEFFZEK, K., VEACH, D., BORNMANN, W., CLARKSON, B., SUPERTI-FURGA, G. & KURIYAN, J. 2003. Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell,* 112**,** 859-71.
NEWMAN, J. R., GHAEMMAGHAMI, S., IHMELS, J., BRESLOW, D. K., NOBLE, M., DERISI, J. L. & WEISSMAN, J. S. 2006. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature,* 441**,** 840-6.
NEWMAN, R. H., HU, J., RHO, H. S., XIE, Z., WOODARD, C., NEISWINGER, J., COOPER, C., SHIRLEY, M., CLARK, H. M., HU, S., HWANG, W., JEONG, J. S., WU, G., LIN, J., GAO, X., NI, Q., GOEL, R., XIA, S., JI, H., DALBY, K. N., BIRNBAUM, M. J., COLE, P. A., KNAPP, S., RYAZANOV, A. G., ZACK, D. J., BLACKSHAW, S., PAWSON, T., GINGRAS, A. C., DESIDERIO, S., PANDEY, A., TURK, B. E., ZHANG, J., ZHU, H. & QIAN, J. 2013. Construction of human activity-based phosphorylation networks. *Mol Syst Biol,* 9**,** 655.
NEWMAN, R. H., ZHANG, J. & ZHU, H. 2014. Toward a systems-level view of dynamic phosphorylation networks. *Front Genet,* 5**,** 263.
NOLEN, B., TAYLOR, S. & GHOSH, G. 2004. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell,* 15**,** 661-75.
NOLLAU, P. & MAYER, B. J. 2001. Profiling the global tyrosine phosphorylation state by Src homology 2 domain binding. *Proc Natl Acad Sci U S A,* 98**,** 13531-6.
O'BRIEN, K. P., REMM, M. & SONNHAMMER, E. L. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res,* 33**,** D476-80.
O'SHEA, J. P., CHOU, M. F., QUADER, S. A., RYAN, J. K., CHURCH, G. M. & SCHWARTZ, D. 2013. pLogo: a probabilistic approach to visualizing sequence motifs. *Nat Methods,* 10**,** 1211-2.
OBENAUER, J. C., CANTLEY, L. C. & YAFFE, M. B. 2003. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res,* 31**,** 3635-41.
ODA, Y., HUANG, K., CROSS, F. R., COWBURN, D. & CHAIT, B. T. 1999. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A,* 96**,** 6591-6.
OH, D., OGIUE-IKEDA, M., JADWIN, J. A., MACHIDA, K., MAYER, B. J. & YU, J. 2012. Fast rebinding increases dwell time of Src homology 2 (SH2)-containing proteins near the plasma membrane. *Proc Natl Acad Sci U S A,* 109**,** 14024-9.
OKADA, M. 2012. Regulation of the SRC family kinases by Csk. *Int J Biol Sci,* 8**,** 1385-97.
OLSEN, J. V., BLAGOEV, B., GNAD, F., MACEK, B., KUMAR, C., MORTENSEN, P. & MANN, M. 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell,* 127**,** 635-48.
PANDYA, S., STRUCK, T. J., MANNAKEE, B. K., PANISCUS, M. & GUTENKUNST, R. N. 2015. Testing whether metazoan tyrosine loss was driven by selection against promiscuous phosphorylation. *Mol Biol Evol,* 32**,** 144-52.
PARRISH, J. R., GULYAS, K. D. & FINLEY, R. L., JR. 2006. Yeast two-hybrid contributions to interactome mapping. *Curr Opin Biotechnol,* 17**,** 387-93.
PARSONS, S. J. & PARSONS, J. T. 2004. Src family kinases, key regulators of signal transduction. *Oncogene,* 23**,** 7906-9.

PASDER, O., SHPUNGIN, S., SALEM, Y., MAKOVSKY, A., VILCHICK, S., MICHAELI, S., MALOVANI, H. & NIR, U. 2006. Downregulation of Fer induces PP1 activation and cell-cycle arrest in malignant cells. *Oncogene,* 25**,** 4194-206.

PATRICK, R., LE CAO, K. A., KOBE, B. & BODEN, M. 2015. PhosphoPICK: modelling cellular context to map kinase-substrate phosphorylation events. *Bioinformatics,* 31**,** 382-9.

PAWSON, T. & NASH, P. 2003. Assembly of cell regulatory systems through protein interaction domains. *Science,* 300**,** 445-52.

PEREZ, O. D. & NOLAN, G. P. 2002. Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. *Nat Biotechnol,* 20**,** 155-62.

PERKINS, D. N., PAPPIN, D. J., CREASY, D. M. & COTTRELL, J. S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis,* 20**,** 3551-67.

PETSALAKI, E., HELBIG, A. O., GOPAL, A., PASCULESCU, A., ROTH, F. P. & PAWSON, T. 2015. SELPHI: correlation-based identification of kinase-associated networks from global phospho-proteomics data sets. *Nucleic Acids Res*.

PICOTTI, P., CLEMENT-ZIZA, M., LAM, H., CAMPBELL, D. S., SCHMIDT, A., DEUTSCH, E. W., ROST, H., SUN, Z., RINNER, O., REITER, L., SHEN, Q., MICHAELSON, J. J., FREI, A., ALBERTI, S., KUSEBAUCH, U., WOLLSCHEID, B., MORITZ, R. L., BEYER, A. & AEBERSOLD, R. 2013. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature,* 494**,** 266-70.

PLUK, H., DOREY, K. & SUPERTI-FURGA, G. 2002. Autoinhibition of c-Abl. *Cell,* 108**,** 247-59.

PRAMATAROVA, A., OCHALSKI, P. G., CHEN, K., GROPMAN, A., MYERS, S., MIN, K. T. & HOWELL, B. W. 2003. Nck beta interacts with tyrosine-phosphorylated disabled 1 and redistributes in Reelin-stimulated neurons. *Mol Cell Biol,* 23**,** 7210-21.

PRASAD, T. S., KANDASAMY, K. & PANDEY, A. 2009. Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods Mol Biol,* 577**,** 67-79.

PTACEK, J., DEVGAN, G., MICHAUD, G., ZHU, H., ZHU, X., FASOLO, J., GUO, H., JONA, G., BREITKREUTZ, A., SOPKO, R., MCCARTNEY, R. R., SCHMIDT, M. C., RACHIDI, N., LEE, S. J., MAH, A. S., MENG, L., STARK, M. J., STERN, D. F., DE VIRGILIO, C., TYERS, M., ANDREWS, B., GERSTEIN, M., SCHWEITZER, B., PREDKI, P. F. & SNYDER, M. 2005. Global analysis of protein phosphorylation in yeast. *Nature,* 438**,** 679-84.

QI, J., WANG, J., ROMANYUK, O. & SIU, C. H. 2006. Involvement of Src family kinases in N-cadherin phosphorylation and beta-catenin dissociation during transendothelial migration of melanoma cells. *Mol Biol Cell,* 17**,** 1261-72.

QIU, H. & MILLER, W. T. 2002. Regulation of the nonreceptor tyrosine kinase Brk by autophosphorylation and by autoinhibition. *J Biol Chem,* 277**,** 34634-41.

RIKOVA, K., GUO, A., ZENG, Q., POSSEMATO, A., YU, J., HAACK, H., NARDONE, J., LEE, K., REEVES, C., LI, Y., HU, Y., TAN, Z., STOKES, M., SULLIVAN, L., MITCHELL, J., WETZEL, R., MACNEILL, J., REN, J. M., YUAN, J., BAKALARSKI, C. E., VILLEN, J., KORNHAUSER, J. M., SMITH, B., LI, D., ZHOU, X., GYGI, S. P., GU, T. L., POLAKIEWICZ, R. D., RUSH, J. & COMB, M. J. 2007. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell,* 131**,** 1190-203.

RITZ, A., SHAKHNAROVICH, G., SALOMON, A. R. & RAPHAEL, B. J. 2009. Discovery of phosphorylation motif mixtures in phosphoproteomics data. *Bioinformatics,* 25**,** 14-21.

ROBINSON, D. R., WU, Y. M. & LIN, S. F. 2000. The protein tyrosine kinase family of the human genome. *Oncogene,* 19**,** 5548-57.

ROSKOSKI, R., JR. 2004. Src protein-tyrosine kinase structure and regulation. *Biochem Biophys Res Commun,* 324**,** 1155-64.

ROSS, A. H., BALTIMORE, D. & EISEN, H. N. 1981. Phosphotyrosine-containing proteins isolated by affinity chromatography with antibodies to a synthetic hapten. *Nature,* 294**,** 654-6.

ROSS, P. L., HUANG, Y. N., MARCHESE, J. N., WILLIAMSON, B., PARKER, K., HATTAN, S., KHAINOVSKI, N., PILLAI, S., DEY, S., DANIELS, S., PURKAYASTHA, S., JUHASZ, P., MARTIN, S., BARTLET-JONES, M., HE, F., JACOBSON, A. & PAPPIN, D. J. 2004. Multiplexed protein quantitation in

Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics,* 3**,** 1154-69.
RUSH, J., MORITZ, A., LEE, K. A., GUO, A., GOSS, V. L., SPEK, E. J., ZHANG, H., ZHA, X. M., POLAKIEWICZ, R. D. & COMB, M. J. 2005. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol,* 23**,** 94-101.
RYBAKIN, V., GOUNKO, N. V., SPATE, K., HONING, S., MAJOUL, I. V., DUDEN, R. & NOEGEL, A. A. 2006. Crn7 interacts with AP-1 and is required for the maintenance of Golgi morphology and protein export from the Golgi. *J Biol Chem,* 281**,** 31070-8.
RYBAKIN, V., RASTETTER, R. H., STUMPF, M., UETRECHT, A. C., BEAR, J. E., NOEGEL, A. A. & CLEMEN, C. S. 2008. Molecular mechanism underlying the association of Coronin-7 with Golgi membranes. *Cell Mol Life Sci,* 65**,** 2419-30.
RYBIN, V. O., GUO, J., GERTSBERG, Z., FEINMARK, S. J. & STEINBERG, S. F. 2008. Phorbol 12-myristate 13-acetate-dependent protein kinase C delta-Tyr311 phosphorylation in cardiomyocyte caveolae. *J Biol Chem,* 283**,** 17777-88.
SAFAEI, J., MANUCH, J., GUPTA, A., STACHO, L. & PELECH, S. 2011. Prediction of 492 human protein kinase substrate specificities. *Proteome Sci,* 9 Suppl 1**,** S6.
SAGAWA, K., SWAIM, W., ZHANG, J., UNSWORTH, E. & SIRAGANIAN, R. P. 1997. Aggregation of the high affinity IgE receptor results in the tyrosine phosphorylation of the surface adhesion protein PECAM-1 (CD31). *J Biol Chem,* 272**,** 13412-8.
SAHA, A., CONNELLY, S., JIANG, J., ZHUANG, S., AMADOR, D. T., PHAN, T., PILZ, R. B. & BOSS, G. R. 2014. Akt phosphorylation and regulation of transketolase is a nodal point for amino acid control of purine synthesis. *Mol Cell,* 55**,** 264-76.
SASAKI, Y., CHENG, C., UCHIDA, Y., NAKAJIMA, O., OHSHIMA, T., YAGI, T., TANIGUCHI, M., NAKAYAMA, T., KISHIDA, R., KUDO, Y., OHNO, S., NAKAMURA, F. & GOSHIMA, Y. 2002. Fyn and Cdk5 mediate semaphorin-3A signaling, which is involved in regulation of dendrite orientation in cerebral cortex. *Neuron,* 35**,** 907-20.
SCHALLER, M. D., HILDEBRAND, J. D., SHANNON, J. D., FOX, J. W., VINES, R. R. & PARSONS, J. T. 1994. Autophosphorylation of the focal adhesion kinase, pp125FAK, directs SH2-dependent binding of pp60src. *Mol Cell Biol,* 14**,** 1680-8.
SCHILLING, O. & OVERALL, C. M. 2008. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol,* 26**,** 685-94.
SCHLESSINGER, J. 2000. Cell signaling by receptor tyrosine kinases. *Cell,* 103**,** 211-25.
SCHLESSINGER, J. 2014. Receptor tyrosine kinases: legacy of the first two decades. *Cold Spring Harb Perspect Biol,* 6.
SCHWARTZ, D., CHOU, M. F. & CHURCH, G. M. 2009. Predicting protein post-translational modifications using meta-analysis of proteome scale data sets. *Mol Cell Proteomics,* 8**,** 365-79.
SCHWARTZ, D. & GYGI, S. P. 2005. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol,* 23**,** 1391-8.
SCHWARTZBERG, P. L., FINKELSTEIN, L. D. & READINGER, J. A. 2005. TEC-family kinases: regulators of T-helper-cell differentiation. *Nat Rev Immunol,* 5**,** 284-95.
SEKIGAWA, M., KUNOH, T., WADA, S., MUKAI, Y., OHSHIMA, K., OHTA, S., GOSHIMA, N., SASAKI, R. & MIZUKAMI, T. 2010. Comprehensive screening of human genes with inhibitory effects on yeast growth and validation of a yeast cell-based system for screening chemicals. *J Biomol Screen,* 15**,** 368-78.
SHEKHAWAT, S. S. & GHOSH, I. 2011. Split-protein systems: beyond binary protein-protein interactions. *Curr Opin Chem Biol,* 15**,** 789-97.
SHETTY, P., VELUSAMY, T., BHANDARY, Y. P., LIU, M. C. & SHETTY, S. 2010. Urokinase receptor expression involves tyrosine phosphorylation of phosphoglycerate kinase. *Mol Cell Biochem,* 335**,** 235-47.
SICHERI, F. & KURIYAN, J. 1997. Structures of Src-family tyrosine kinases. *Curr Opin Struct Biol,* 7**,** 777-85.

SING, T., SANDER, O., BEERENWINKEL, N. & LENGAUER, T. 2005. ROCR: visualizing classifier performance in R. *Bioinformatics,* 21**,** 3940-1.
SINGER, D., KUHLMANN, J., MUSCHKET, M. & HOFFMANN, R. 2010. Separation of multiphosphorylated peptide isomers by hydrophilic interaction chromatography on an aminopropyl phase. *Anal Chem,* 82**,** 6409-14.
SIRAGANIAN, R. P., ZHANG, J., SUZUKI, K. & SADA, K. 2002. Protein tyrosine kinase Syk in mast cell signaling. *Mol Immunol,* 38**,** 1229-33.
SMOLKA, M. B., ALBUQUERQUE, C. P., CHEN, S. H. & ZHOU, H. 2007. Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proc Natl Acad Sci U S A,* 104**,** 10364-9.
SNEL, B., LEHMANN, G., BORK, P. & HUYNEN, M. A. 2000. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res,* 28**,** 3442-4.
SON, J., LYSSIOTIS, C. A., YING, H., WANG, X., HUA, S., LIGORIO, M., PERERA, R. M., FERRONE, C. R., MULLARKY, E., SHYH-CHANG, N., KANG, Y., FLEMING, J. B., BARDEESY, N., ASARA, J. M., HAIGIS, M. C., DEPINHO, R. A., CANTLEY, L. C. & KIMMELMAN, A. C. 2013. Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. *Nature,* 496**,** 101-5.
SONGYANG, Z., BLECHNER, S., HOAGLAND, N., HOEKSTRA, M. F., PIWNICA-WORMS, H. & CANTLEY, L. C. 1994. Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr Biol,* 4**,** 973-82.
SONGYANG, Z. & CANTLEY, L. C. 1995. Recognition and specificity in protein tyrosine kinase-mediated signalling. *Trends Biochem Sci,* 20**,** 470-5.
SOPKO, R. & ANDREWS, B. J. 2008. Linking the kinome and phosphorylome--a comprehensive review of approaches to find kinase targets. *Mol Biosyst,* 4**,** 920-33.
SOPKO, R., FOOS, M., VINAYAGAM, A., ZHAI, B., BINARI, R., HU, Y., RANDKLEV, S., PERKINS, L. A., GYGI, S. P. & PERRIMON, N. 2014. Combining genetic perturbations and proteomics to examine kinase-phosphatase networks in Drosophila embryos. *Dev Cell,* 31**,** 114-27.
SRIRAM, G., JANKOWSKI, W., KASIKARA, C., REICHMAN, C., SALEH, T., NGUYEN, K. Q., LI, J., HORNBECK, P., MACHIDA, K., LIU, T., LI, H., KALODIMOS, C. G. & BIRGE, R. B. 2014. Iterative tyrosine phosphorylation controls non-canonical domain utilization in Crk. *Oncogene*.
SRIRAM, G., REICHMAN, C., TUNCEROGLU, A., KAUSHAL, N., SALEH, T., MACHIDA, K., MAYER, B., GE, Q., LI, J., HORNBECK, P., KALODIMOS, C. G. & BIRGE, R. B. 2011. Phosphorylation of Crk on tyrosine 251 in the RT loop of the SH3C domain promotes Abl kinase transactivation. *Oncogene,* 30**,** 4645-55.
STELZL, U., WORM, U., LALOWSKI, M., HAENIG, C., BREMBECK, F. H., GOEHLER, H., STROEDICKE, M., ZENKNER, M., SCHOENHERR, A., KOEPPEN, S., TIMM, J., MINTZLAFF, S., ABRAHAM, C., BOCK, N., KIETZMANN, S., GOEDDE, A., TOKSOZ, E., DROEGE, A., KROBITSCH, S., KORN, B., BIRCHMEIER, W., LEHRACH, H. & WANKER, E. E. 2005. A human protein-protein interaction network: a resource for annotating the proteome. *Cell,* 122**,** 957-68.
STURM, M., BERTSCH, A., GROPL, C., HILDEBRANDT, A., HUSSONG, R., LANGE, E., PFEIFER, N., SCHULZ-TRIEGLAFF, O., ZERCK, A., REINERT, K. & KOHLBACHER, O. 2008. OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinformatics,* 9**,** 163.
SUGA, H., DACRE, M., DE MENDOZA, A., SHALCHIAN-TABRIZI, K., MANNING, G. & RUIZ-TRILLO, I. 2012. Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. *Sci Signal,* 5**,** ra35.
SULLIVAN, R. & GRAHAM, C. H. 2007. Hypoxia-driven selection of the metastatic phenotype. *Cancer Metastasis Rev,* 26**,** 319-31.
SUPEK, F., BOSNJAK, M., SKUNCA, N. & SMUC, T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One,* 6**,** e21800.
SUPERTI-FURGA, G., FUMAGALLI, S., KOEGL, M., COURTNEIDGE, S. A. & DRAETTA, G. 1993. Csk inhibition of c-Src activity requires both the SH2 and SH3 domains of Src. *EMBO J,* 12**,** 2625-34.

SYLVESTER, M., KLICHE, S., LANGE, S., GEITHNER, S., KLEMM, C., SCHLOSSER, A., GROSSMANN, A., STELZL, U., SCHRAVEN, B., KRAUSE, E. & FREUND, C. 2010. Adhesion and degranulation promoting adapter protein (ADAP) is a central hub for phosphotyrosine-mediated interactions in T cells. *PLoS One,* 5**,** e11708.

TAKASHIMA, Y., DELFINO, F. J., ENGEN, J. R., SUPERTI-FURGA, G. & SMITHGALL, T. E. 2003. Regulation of c-Fes tyrosine kinase activity by coiled-coil and SH2 domains: analysis with Saccharomyces cerevisiae. *Biochemistry,* 42**,** 3567-74.

TAN, C. S., BODENMILLER, B., PASCULESCU, A., JOVANOVIC, M., HENGARTNER, M. O., JORGENSEN, C., BADER, G. D., AEBERSOLD, R., PAWSON, T. & LINDING, R. 2009a. Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci Signal,* 2**,** ra39.

TAN, C. S., PASCULESCU, A., LIM, W. A., PAWSON, T., BADER, G. D. & LINDING, R. 2009b. Positive selection of tyrosine loss in metazoan evolution. *Science,* 325**,** 1686-8.

TAN, J. L. & SPUDICH, J. A. 1990. Developmentally regulated protein-tyrosine kinase genes in Dictyostelium discoideum. *Mol Cell Biol,* 10**,** 3578-83.

TINTI, M., NARDOZZA, A. P., FERRARI, E., SACCO, F., CORALLINO, S., CASTAGNOLI, L. & CESARENI, G. 2012. The 4G10, pY20 and p-TYR-100 antibody specificity: profiling by peptide microarrays. *N Biotechnol,* 29**,** 571-7.

TISDALE, E. J. & ARTALEJO, C. R. 2006. Src-dependent aprotein kinase C iota/lambda (aPKCiota/lambda) tyrosine phosphorylation is required for aPKCiota/lambda association with Rab2 and glyceraldehyde-3-phosphate dehydrogenase on pre-golgi intermediates. *J Biol Chem,* 281**,** 8436-42.

TISDALE, E. J. & ARTALEJO, C. R. 2007. A GAPDH mutant defective in Src-dependent tyrosine phosphorylation impedes Rab2-mediated events. *Traffic,* 8**,** 733-41.

TRISTAN, C., SHAHANI, N., SEDLAK, T. W. & SAWA, A. 2011. The diverse functions of GAPDH: views from different subcellular compartments. *Cell Signal,* 23**,** 317-23.

UBERSAX, J. A. & FERRELL, J. E., JR. 2007. Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol,* 8**,** 530-41.

UMMANNI, R., MANNSPERGER, H. A., SONNTAG, J., OSWALD, M., SHARMA, A. K., KONIG, R. & KORF, U. 2014. Evaluation of reverse phase protein array (RPPA)-based pathway-activation profiling in 84 non-small cell lung cancer (NSCLC) cell lines as platform for cancer proteomics and biomarker discovery. *Biochim Biophys Acta,* 1844**,** 950-9.

UNIPROT, C. 2015. UniProt: a hub for protein information. *Nucleic Acids Res,* 43**,** D204-12.

VANDER HEIDEN, M. G., CANTLEY, L. C. & THOMPSON, C. B. 2009. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science,* 324**,** 1029-33.

VARJOSALO, M., KESKITALO, S., VAN DROGEN, A., NURKKALA, H., VICHALKOVSKI, A., AEBERSOLD, R. & GSTAIGER, M. 2013. The protein interaction landscape of the human CMGC kinase group. *Cell Rep,* 3**,** 1306-20.

VILA, J. M., GIMFERRER, I., PADILLA, O., ARMAN, M., PLACES, L., SIMARRO, M., VIVES, J. & LOZANO, F. 2001. Residues Y429 and Y463 of the human CD5 are targeted by protein tyrosine kinases. *Eur J Immunol,* 31**,** 1191-8.

VINAYAGAM, A., HU, Y., KULKARNI, M., ROESEL, C., SOPKO, R., MOHR, S. E. & PERRIMON, N. 2013. Protein complex-based analysis framework for high-throughput data sets. *Sci Signal,* 6**,** rs5.

WAGIH, O., REIMAND, J. & BADER, G. D. 2015. MIMP: predicting the impact of mutations on kinase-substrate phosphorylation. *Nat Methods*.

WANG, C., YE, M., BIAN, Y., LIU, F., CHENG, K., DONG, M., DONG, J. & ZOU, H. 2013. Determination of CK2 specificity and substrates by proteome-derived peptide libraries. *J Proteome Res,* 12**,** 3813-21.

WANG, J., WANG, J., DAI, J., JUNG, Y., WEI, C. L., WANG, Y., HAVENS, A. M., HOGG, P. J., KELLER, E. T., PIENTA, K. J., NOR, J. E., WANG, C. Y. & TAICHMAN, R. S. 2007. A glycolytic mechanism regulating an angiogenic switch in prostate cancer. *Cancer Res,* 67**,** 149-59.

WARBURG, O. 1956. On the origin of cancer cells. *Science,* 123**,** 309-14.

WARD, J. J., MCGUFFIN, L. J., BRYSON, K., BUXTON, B. F. & JONES, D. T. 2004. The DISOPRED server for the prediction of protein disorder. *Bioinformatics,* 20**,** 2138-9.
WATTS, D. J. & STROGATZ, S. H. 1998. Collective dynamics of 'small-world' networks. *Nature,* 393**,** 440-2.
WEIJLAND, A., WILLIAMS, J. C., NEUBAUER, G., COURTNEIDGE, S. A., WIERENGA, R. K. & SUPERTI-FURGA, G. 1997. Src regulated by C-terminal phosphorylation is monomeric. *Proc Natl Acad Sci U S A,* 94**,** 3590-5.
WEPF, A., GLATTER, T., SCHMIDT, A., AEBERSOLD, R. & GSTAIGER, M. 2009. Quantitative interaction proteomics using mass spectrometry. *Nat Methods,* 6**,** 203-5.
WESTHEIMER, F. H. 1987. Why nature chose phosphates. *Science,* 235**,** 1173-8.
WOODARD, C. L., GOODWIN, C. R., WAN, J., XIA, S., NEWMAN, R., HU, J., ZHANG, J., HAYWARD, S. D., QIAN, J., LATERRA, J. & ZHU, H. 2013. Profiling the dynamics of a human phosphorylome reveals new components in HGF/c-Met signaling. *PLoS One,* 8**,** e72671.
WOODSMITH, J., KAMBUROV, A. & STELZL, U. 2013. Dual coordination of post translational modifications in human protein networks. *PLoS Comput Biol,* 9**,** e1002933.
WORSECK, J. M., GROSSMANN, A., WEIMANN, M., HEGELE, A. & STELZL, U. 2012. A stringent yeast two-hybrid matrix screening approach for protein-protein interaction discovery. *Methods Mol Biol,* 812**,** 63-87.
XUE, L., WANG, W. H., ILIUK, A., HU, L., GALAN, J. A., YU, S., HANS, M., GEAHLEN, R. L. & TAO, W. A. 2012. Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. *Proc Natl Acad Sci U S A,* 109**,** 5615-20.
XUE, Y., LI, A., WANG, L., FENG, H. & YAO, X. 2006. PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinformatics,* 7**,** 163.
XUE, Y., REN, J., GAO, X., JIN, C., WEN, L. & YAO, X. 2008. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol Cell Proteomics,* 7**,** 1598-608.
YAMASHITA, T., SUZUKI, R., BACKLUND, P. S., YAMASHITA, Y., YERGEY, A. L. & RIVERA, J. 2008. Differential dephosphorylation of the FcRgamma immunoreceptor tyrosine-based activation motif tyrosines with dissimilar potential for activating Syk. *J Biol Chem,* 283**,** 28584-94.
YANG, J., KIM, O., WU, J. & QIU, Y. 2002. Interaction between tyrosine kinase Etk and a RUN domain- and FYVE domain-containing protein RUFY1. A possible role of ETK in regulation of vesicle trafficking. *J Biol Chem,* 277**,** 30219-26.
YANG, X., HUBBARD, E. J. & CARLSON, M. 1992. A protein kinase substrate identified by the two-hybrid system. *Science,* 257**,** 680-2.
ZHU, H., BILGIN, M., BANGHAM, R., HALL, D., CASAMAYOR, A., BERTONE, P., LAN, N., JANSEN, R., BIDLINGMAIER, S., HOUFEK, T., MITCHELL, T., MILLER, P., DEAN, R. A., GERSTEIN, M. & SNYDER, M. 2001. Global analysis of protein activities using proteome chips. *Science,* 293**,** 2101-5.
ZIPFEL, P. A., ZHANG, W., QUIROZ, M. & PENDERGAST, A. M. 2004. Requirement for Abl kinases in T cell receptor signaling. *Curr Biol,* 14**,** 1222-31.

Acknowledgements

First of all, I want to thank Dr. Ulrich Stelzl for excellent supervision during the entire time of my doctoral studies. He taught me how to tackle scientific questions while always being available for simplest questions and fruitful discussions. I furthermore want to thank all former and present members of the laboratory, Nouhad Benlasfer, Anna Hegele, Petra Birth, Josefine Worseck, Mareike Weimann, Arndt Grossmann, Stefanie Jehle, and Luise Apelt, for patiently helping me with all the small and big problems I faced. In particular, I want to thank Dr. Jonathan Woodsmith who helped me a lot by running his algorithm on my data, discussions about the project, reading my thesis, and by continuing work on a second project I had started. Moreover, I want to thank Federico Apelt, a former Master student whom I supervised and who supervised me when I started with python programming and who performed data filtering and normalization. Furthermore, I thank Bryan Ballif (University of Vermont, USA) and David Meierhofer (MPI-MG) and his lab who both measured all my mass spectrometry samples and provided great advice how to conduct the experiments. Thanks also to Sean R. Connell (JWGU, Frankfurt) for mapping our data on yeast complex structures and to Miguel Andrade (MDC, Berlin) and Jean-Fred Fontaine (MDC, Berlin) for their help with the phylogenic analysis and discussions. I also want to thank Johannes Helmuth (MPI-MG) for bioinformatics support and all other students at the MPI-MG and the FU Berlin I worked with. Finally, I want to thank my wife Miganoush Magarian and our families for supporting me in general.

Publications

HINKLE, K., WEIR, M., FULTON, Z., HAO, J., MANN, J., MCGEHEE, A., CORWIN, T., STELZL, U., DEMING, P., JUO, P. & BALLIF, B. 2015. Novel Tyrosine Phosphorylation Sites Fine Tune the Activity and Substrate Binding of Src Family Kinases. *The FASEB Journal,* 29.

Curriculum Vitae

For reasons of data protection,

the curriculum vitae is not included in the online version