Systems biological analysis of the platelet proteome and applications of functional module search in proteome networks

***

Systembiologische Analyse des Blutplättchenproteoms und funktionelle Modulsuche in Proteinnetzwerken

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrades

der Julius-Maximilians-Universität Würzburg

vorgelegt von

Desislava Veselinova Boyanova

Geburtsort: Plovdiv, Bulgarien

Würzburg 2012

Eingereicht am:

....

Mitglieder der Promotionskommission:

Vorsitzender: ....

Erster Gutachter: Prof. Dr. Thomas Dandekar

Zweiter Gutachter: Prof. Dr. Markus Engstler

Tag des Promotionskolloquiums:……………………………………………………….

Doktorurkunde ausgehändigt am:………………………………………………………

Affidavit / Eidesstattliche Erklärung

English:

I hereby confirm that my thesis entitled "Systems biological analysis of the platelet proteome and applications of functional module search in proteome networks" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Additionally, other than this degree, I have not applied or will attempt to apply for any other degree or qualification in relation to this thesis.

Würzburg, Signature:

Place, Date (Desislava Boyanova)

Deutsch:

Hiermit erkläre ich an Eides statt, die Dissertation „Systembiologische Analyse des Blutplättchenproteoms und funktionelle Modulsuche in Proteinnetzwerken“ eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Zusätzlich habe oder werde ich nicht versuchen neben diesem Abschluss einen weiteren Abschluss oder Qualifikation mit dieser Doktorarbeit zu erwerben.

Würzburg, Unterschrift:

Ort, Datum (Desislava Boyanova)

Acknowledgements

This work was accomplished with the help and support of a number of people. First, I want to thank my supervisor and doctor father Prof. Thomas Dandekar for his guidance and support throughout my PhD thesis and for introducing me to all exciting fields of bioinformatics. My gratitude also goes to my second supervisor Prof. Markus Engstler for kindly reviewing this thesis. I am grateful to my PhD supervisor Dr. Dr. Marcus Dittrich for his fruitful discussions, valuable lessons and useful advice in every aspect of this dissertation. Dr. Tobias Müller also greatly influenced my work by adding important aspects of statistics and invaluable suggestions for the optimization of the module detection algorithm.

I appreciate the help of Prof. Jörg Schultz and all my colleagues from the literature seminar meeting for critically evaluating many of the results presented here and helping me to improve my scientific thinking. My colleague, Santosh Nilla, has been a big inspiration to me, thank you for going through all the rough patches of these scientific challenges with me, for always persisting with unwavering optimism and showing me how collaborative work achieves great goals! I also want to thank Gaby Wangorsch, the other true „plateleteer”, who always gave me a helping hand in the mysterious paths of platelet signaling, Daniela Beisser for her invaluable insights on the functional module detection algorithm and statistical background and Benjamin Schütte for his honest critical remarks on all my results presented in internal meetings. My gratitude goes to Astrid Fieselmann and Alexander Cecil for their support and advice on my PhD thesis. Last but not least, I want to thank Christian Koetschan for his help in every aspect, from advice on the optimization of Perl scripts and database issues to moral support and true inspiration for the writing and finalization of this PhD thesis. To the best colleagues one can hope to have – thank you for all the exciting and fun days and for making me come to work with a smile on my face! I would also like to thank our collaboration partners, who made this project possible in the first place: the clinical biochemistry groups (Dr. Jörg Geiger, Dr. Stepan Gambaryan and Prof. Ulrich Walter) for performing all experiments with platelets and the proteomics group of Prof. Sickmann for the qualitative and quantitative mass spectrometry data.

My final appreciation goes to my family and their unwavering support and patience, for giving me the will and courage to fulfill my scientific aspirations. To them and to my closest friends Desi, Louiza and Ani, who stood beside me every step of the long way leading to this page – thank you for always being there!

Table of Contents

1 Abstract

Recent development of proteomic approaches and generation of large-scale proteomic datasets calls for new methods for biological interpretation of the obtained results. Systems biological approaches such as integrated network analysis and functional module search have become an essential part of proteomic investigation. Proteomics is especially applied in anucleate cells such as platelets. The underlying molecular mechanisms of platelet activation and their pharmacological modulation are of immense importance for clinical research. Advances in platelet proteomics have provided a large amount of proteomic data, which has not yet been comprehensively investigated in a systems biological perspective.

To this end, I assembled platelet specific data from proteomic and transcriptomic studies by detailed manual curation and worked on the generation of a comprehensive human platelet repository for systems biological analysis of platelets in the functional context of integrated networks (*PlateletWeb*) (*http:/PlateletWeb.bioapps.biozentrum.uni-wuerzburg.de)*. I also added platelet-specific experimentally validated phosphorylation data and generated kinase predictions for 80% of the newly identified platelet phosphosites. The combination of drug, disease and pathway information with phosphorylation and interaction data makes this database the first integrative platelet platform available for platelet research. *PlateletWeb* contains more than 5000 platelet proteins, which can also be analyzed and visualized in a network context, allowing identification of all major signaling modules involved in platelet activation and inhibition.

Using the wealth of integrated data I performed a series of platelet-specific analyses regarding the platelet proteome, pathways, drug targets and novel platelet phosphorylation events involved in crucial signaling events. I analyzed the statistical enrichment of known pathways for platelet proteins and identified endocytosis as a highly represented pathway in platelets. Further results revealed that highly connected platelet proteins are more often targeted by drugs.

Using integrated network analysis offered by *PlateletWeb*, I analyzed the crucial activation signaling pathway of adenosine diphosphate (ADP), visualizing how the signal flow from receptors to effectors is maintained. My work on integrin inside-out signaling was also based on the integrated network approach and examined new platelet-specific phosphorylation sites and their regulation using kinase predictions. I generated hypothesis on integrin signaling, by investigating the regulation of Ser^{269} phosphorylation site on the docking protein 1 (DOK1). This phosphorylation site may influence the inhibiting effect of DOK1 on integrin α2bβ3.

Extending the integrated network approach to further cell lines, I used the assembled human interactome information for the analysis of functional modules in cellular networks. The investigation was performed with a previously developed module detection algorithm, which finds maximum-scoring subgraphs in transcriptomic datasets by using assigned values to the network nodes. We extended the algorithm to qualitative proteomic datasets and enhanced the module search by adding functional information to the network edges to concentrate the solution onto modules with high functional similarity. I performed a series of analyses to validate its performance in small-sized (virus-infected gastric cells) and medium-sized networks (human lymphocytes). In both cases the algorithm extracted characteristic modules of sample proteins with high functional similarity.

The functional module search is especially useful in site-specific phosphoproteomic datasets, where kinase regulation of the detected sites is often sparse or lacking. Therefore, I used the module detection algorithm in quantitative phosphoproteomic datasets. In a platelet phosphorylation dataset, I presented a pipeline for network analysis of detected phosphorylation sites. In a second approach, the functional module detecting algorithm was used on a phosphoproteome network of human embryonic stem cells, in which nodes represented the maximally changing phosphorylation sites in the experiment. Additional kinases from the human phosphoproteome in *PlateletWeb* were included to the network to investigate the regulation of the signal flow. Results indicated important phosphorylation sites and their upstream kinases and explained changes observed in embryonic stem cells during differentiation.

This work presents novel approaches for integrated network analysis in cells and introduces for the first time a systematic biological investigation of the human platelet proteome based on the platelet-specific knowledge base *PlateletWeb*. The extended methods for optimized functional module detection offer an invaluable tool for exploring proteomic datasets and covering gaps in complex large-scale data analysis. By combining exact module detection approaches with functional information data between interacting proteins, characteristic functional modules with high functional resemblance can be extracted from complex datasets, thereby focusing on important changes in the observed networks.

2 Zusammenfassung

Jüngste Entwicklungen der Proteomik und die damit einhergehende Erzeugung großer Datensätze erfordern neue Methoden zur biologischen Interpretation der gewonnenen Ergebnisse. Systembiologische Ansätze wie die integrierte Netzwerkanalyse sowie die funktionelle Modulsuche sind zu einem wesentlichen Bestandteil bei der Untersuchung von Proteinen geworden. Die Proteomik wird vor allem in kernlosen Zellen wie den Blutplättchen angewandt. Die zu Grunde liegenden molekularen Mechanismen bei der Aktivierung von Thrombozyten und deren pharmakologische Modulation sind von immenser Bedeutung für die klinische Forschung. Aktuelle Studien in der Proteomforschung haben insbesondere bei Thrombozyten große Mengen an Daten erzeugt, die bisher noch nicht umfassend systembiologisch untersucht wurden.

Zu diesem Zweck stellte ich manuell thrombozyten-spezifische Daten aus Proteom- und Transkriptomstudien zusammen und arbeitete an der Entwicklung einer umfassenden menschlichen Thrombozytendatenbank für die systembiologische Analyse der Funktion von Blutplättchen mittels integrierter Netzwerkanalyse (*PlateletWeb*) (*http:/PlateletWeb.bioapps.biozentrum.uni-wuerzburg.de*). Zusätzlich habe ich plättchen-spezifische, experimentell validierte Phosphorylierungsinformationen hinzugefügt und generierte Kinasenvorhersagen für 80% der neu identifizierten Phosphorylierungsstellen. Die Kombination aus Medikamenten, assoziierten Krankheiten und Signalweginformation zusammen mit Phosphorylierungs- und Interaktionsdaten macht diese Datenbank zu einer ersten und umfassenden Anlaufstelle für Thrombozytenforschung. *PlateletWeb* enthält mehr als 5000 Plättchenproteine, die in einem Netzwerk analysiert und dargestellt werden können. Dabei ist die Identifizierung aller wichtigen Signalmodule zur Plättchenaktivierung und -inhibierung möglich.

Mit der Fülle an verfügbaren Daten führte ich eine Reihe thrombozyten-spezifischer Analysen am Plättchenproteom, an Signalwegen, pharmakologischen Wirkstoffzielen und Phosphorylierungsreaktionen in grundlegenden Signalprozessen durch. Ich analysierte die statistische Anreicherung bekannter Signalwege für Plättchenproteine und identifizierte Endozytose als einen sehr repräsentativen Signalweg in Thrombozyten. Weitere Ergebnisse zeigten, dass stark vernetzte Plättchenproteine häufiger Ziel von Medikamenten sind.

Mittels der Netzwerkanalyse von *PlateletWeb* untersuchte ich den grundlegenden Signalaktivierungspfad von Adenosindiphosphat (ADP), und veranschaulichte den Signalfluss von Rezeptor zu Effektor. Meine Arbeit an der Integrin-Inside-Out-Signalisierung beinhaltete zudem die Untersuchung neuer thrombozyten-spezifischer Phosphorylierungsstellen und ihre Regulation durch Kinasenvorhersagen mit Hilfe des integrierten Netzwerkanalyseansatzes. Durch die Untersuchung der Regulation bei der Phosphorylierungsstelle Ser^{269} im Docking-Protein (DOK1) stellte ich eine neue Hypothese zur Integrinsignalisierung auf. Diese Phosphorylierungsstelle könnte den inhibitorischen Effekt von DOK1 auf integrin α2bβ3 beeinflussen.

Ich erweiterte den integrierten Netzwerkanalyseansatz für andere Zelllinien, indem ich die gesammelten Informationen aus dem menschlichen Interaktom für die Analyse von funktionellen Modulen in zellulären Netzen nutzte. Die Untersuchung wurde mit einem zuvor entwickelten Algorithmus zur Modulerkennung durchgeführt, der maximal bewertete Teilgraphen in Transkriptomdatensätzen anhand zugewiesener Werte für Netzwerkknoten findet. Wir erweiterten den Algorithmus zur Anwendung auf qualitative Proteomdatensätze und optimierten die Modulsuche durch Integration funktioneller Informationen in die Netzwerkkanten. Dies fokussierte die Optimierung auf Proteinmodule mit hoher funktioneller Ähnlichkeit. Ich führte eine Reihe von Analysen durch, um die Effizienz des Algorithmus in kleinen (durch Viren infizierte Magenzellen) und mittelgroßen Netzwerken (menschliche Lymphozyten) zu überprüfen. In beiden Fällen extrahierte der Algorithmus charakteristische Module der untersuchten Proteine mit hohen funktionellen Ähnlichkeiten.

Die funktionelle Modulsuche ist besonders bei positionsspezifischen Phosphoproteomikdatensätzen nützlich, in denen die Kinasenregulation der detektierten Phosphorylierungsstellen nur spärlich oder gar nicht vorhanden ist. Daher habe ich den Algorithmus der Moduldetektion auf quantitative Phosphoproteomikdatensätze angewandt. Anhand eines Datensatzes bestehend aus phosphorylierten Plättchenproteinen habe ich eine Vorgehensweise zur Netzwerkanalyse von Phosphorylierungsstellen entwickelt. In einer zweiten Studie wurde der Algorithmus der Moduldetektion auf ein phosphoproteomisches Netzwerk menschlich embryonaler Stammzellen angewandt, in dem Phosphorylierungsstellen mit maximaler Veränderung durch Netzwerkknoten repräsentiert wurden. Um die Regulation des Signalflusses zu untersuchen wurden weitere Kinasen aus dem menschlichen Phosphoproteom beziehungsweise *PlateletWeb* integriert. Ergebnisse wiesen auf wichtige

Phosphorylierungsstellen und ihre Upstream-Kinasen hin und verdeutlichten Vorgänge, die während der Differenzierung in den embryonalen Stammzellen stattgefunden haben.

Diese Arbeit bietet neue Vorgehensweisen der integrierten Netzwerkanalyse in Zellen und präsentiert zum ersten Mal eine systembiologische Untersuchung des menschlichen Proteoms mit Hilfe der Trombozytendatenbank *PlateletWeb*. Die erweiterten Methoden zur verbesserten Erkennung funktioneller Module bieten ein wertvolles Werkzeug für die Erforschung proteomischer Datensätze und vervollständigen die komplexe und umfangreiche Datenanalyse.

Charakteristische Module, die große Ähnlichkeit auf funktioneller Ebene aufweisen, können durch die Kombination von exakten Modulerkennungsansätzen mit funktionellen Daten extrahiert werden. Dabei werden wichtige Änderungen besonders bei der Analyse komplexer Netzwerke hervorgehoben.

3 Introduction

3.1 The role of proteomics and phosphoproteomics in systems biology

The most advanced research arises from fundamental questions: How is a cell defined? How does one analyze cell signaling? These central issues are the main focus of systems biology today. Every biological entity is a complex system of a variety of processes, each coupled to one another in an admirable harmony. When this harmony is disrupted defects appear causing the outbreak of disease. Medicine has fought for years to understand the biological mechanisms of disease, yet full comprehension was achieved mainly on genetic diseases and diseases associated with a single dysfunction. Complex diseases with multiple causes such as diabetes and atherosclerosis are only to be interpreted in the context of systems biology, as there are often multiple factors and imbalances causing these diseases (Dempfle, Scherag et al. 2008).It has been increasingly recognized in recent years that the understanding of changes and properties that arise from whole-cell function require integrated analysis of the relationships between different cellular components (Albert 2005). To deepen our understanding in the complex cellular mechanisms one cannot merely focus on the single-protein or receptor level (reductionist view) but should rather concentrate on the cell system as a whole (wholist view)(Junker and Schreiber 2008). The entire set of components with a particular characteristic is described with terms ending with “ome” (genome, proteome), while the techniques to identify the set acquires the ending “omics” (genomics, proteomics). In this sense, the term “proteome” represents the set of all proteins expressed and measured in a cell and from it the word interactome is derived to illustrate all physical interactions found in a cell. The word "interactome" was originally introduced in 1999 by a group of French scientists headed by Bernard Jacq (Sanchez, Lachaize et al. 1999). Further fields also developed presenting more characteristics such as metabolomics (Fiehn, Kopka et al. 2000) and the human diseasome (Goh, Cusick et al. 2007) etc.

The central change in systems biology has been the switch from bottom-up to top-down approaches (Katagiri 2003). The beginning of systems biology as a field has been marked by different efforts. In the first half of the 20th century, enzyme kinetics focused only on kinetics and interactions. Then, Ludwig von Bertalanffy proposed his systems theory in 1968 (Bertalanffy 1968). In the following years the interest for this field has increased immensely. Hiroaki Kitano defined systems biology in his book “Foundations of Systems Biology” as “systems biology is a new field in biology that aims at system-level understanding of biological systems” (Kitano 2001). Subsequently, a whole new world has been revealed paving the way for systems biologists to investigate the intricate signaling events in human cells. Systems biological approaches gain importance, because reductionists’ methods lack the needed complexity to explain changes observed in cell behavior under a predefined set of conditions. Beginning from the sequencing of the human genome in 2001 (Lander, Linton et al. 2001; Venter, Adams et al. 2001), the genomics field flourished along with technical advances and computational power consequently solving the previously unthinkable task of creating a guiding map to the human gene repertoire. Thus, a first big collaborative step was made towards decoding the complicated processes hidden in human cells and a new way of thinking was introduced to scientists, demanding a whole array of novel approaches to solve the newly arisen challenges. Genome investigation led to the realization that the set of mRNAs available in a cell (transcriptome) can be used to create a first glimpse of the changes induced by external stimuli. The dynamic changes of the transcriptome became the focus of a whole new line of techniques, starting with the microarray (McGall and Christians 2002) and the subsequently developed RNA sequencing (Wang, Gerstein et al. 2009). Nonetheless, RNA and corresponding protein levels have been known to correlate poorly (Pascal, True et al. 2008; Olsen, Vermeulen et al. 2010). Soon enough the realization came that objective comprehension of what is really happening in a cell can be achieved only by measuring the protein expression at a given time under specific conditions.

The proteomics era began, opening a wave of opportunities to track, analyze and interpret cell responses. The development of proteomics was facilitated by a huge leap in the field of mass spectrometry. The basic principle of mass spectrometry analysis lies in the digestion of a protein probe and consequent ionization of the obtained peptides using an ionization source. Due to their positive charge, the ionized peptides can be accelerated using an electrical field in the mass spectrometer. Each ion is detected according to the mass-to-charge ratio and identified peptides are finally mapped to known proteins using bioinformatical approaches and software. Proteome studies have become an established and widely used method for the analysis of protein samples and investigation of signaling pathways under different conditions (Preisinger, von Kriegsheim et al. 2008; Choudhary and Mann 2010). Various strategies have been developed to investigate different aspects of the cell proteome. The most widespread methods for analysis of global protein expression are based on two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) allowing the resolution of large proteomic samples comprised of thousands of proteins. High precision and accuracy can be achieved by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), which has been established as a high-throughput technique (Rotilio, Della Corte et al. 2012).

Soon it was not sufficient only to measure the presence or absence of proteins in a particular probe, but also their posttranslational modifications (PTMs). These alterations in protein structure are caused by adding or removing chemical groups and induce a change in the tertiary protein structure triggering changes in the protein function. One such modification, which plays a key role in cell signaling, is protein phosphorylation. During phosphorylation a special enzyme called “kinase”, adds a phosphate group to a protein, thus either activating or inhibiting its main biological function. The counteracting enzyme, called “phosphatase”, removes the added phosphorylation group to stop the induced signal (Figure 1).

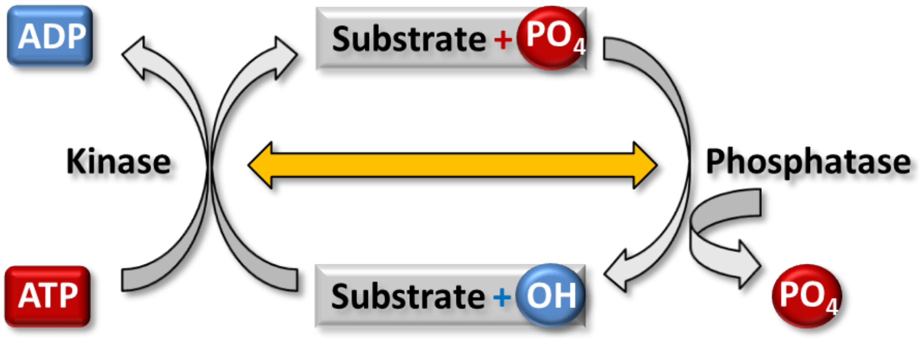


Figure 1. Phosphorylation and dephosphorylation

Kinases and phosphatases are counterplayers in signaling events. Kinases add a phosphate group (PO_4) extracted from adenosine triphosphate (ATP) to a Serine, Threonine or Tyrosine residue of the substrate, which are then removed by phosphatases.

It is believed that one-thirds of the eukaryotic proteome is modified by phosphorylation of mainly serine, threonine or tyrosine residues (Mann, Ong et al. 2002). Phosphorylation is one of the best characterized protein modifications due to its participation in signal transduction pathways, which have been thoroughly analyzed biochemically. In some anucleate cell types, such as platelets, protein modifications and especially phosphorylation are critically involved in maintaining cellular functions and responses. Once proteins in a single probe could be identified along with their PTMs, the question arose whether one could measure these changes precisely in the concept of abundance. Thus, quantitative proteomics methods paved the way for a new understanding of cellular functions based on quantitative data. The most common quantitation methods are label-free approaches or isotopic labelling methods: i) stable isotope labelling by amino acids in cell culture (SILAC, metabolic labelling)(Mann, Ong et al. 2002) and ii) isobaric tags for relative and absolute quantitation (ITRAQ, chemical labeling)(Barnouin 2012). Quantitative analysis has been developed in a series of cell lines (Brill, Xiong et al. 2009), investigating cellular subcompartments (Malik, Lenobel et al. 2009) and various cellular conditions (Daub, Olsen et al. 2008; de Godoy, Olsen et al. 2008; Boisvert, Lam et al. 2010; Zhong, Krawczyk et al. 2010). The impact of quantitative proteomics in systems biology and in cellular signaling analysis in particular, has been immense as measuring of protein abundance is a direct marker for changes on the cellular level and can improve the understanding of complex network dynamics (Preisinger, von Kriegsheim et al. 2008; White 2008; Jorgensen and Locard-Paulet 2012).

3.2 Integrated network analysis

A network in biological sense is a simplified model that describes a set of molecules, e.g. proteins or genes, which are connected with each other through well-defined relationships. Graph properties allow the investigation of complex biological interactions using mathematical and computational approaches through simplifying the representation of these interactions in a network context to reveal regulation patterns often difficult to measure experimentally. The various application fields of networks include the representation of metabolites in metabolic networks, gene regulation in regulatory networks, protein-protein interactions in PPI networks, genetic interactions in gene networks or visualization of Gene Ontology terms of proteins in a directed acyclic graph (DAG). Network properties of biological networks share a lot of common properties with other natural or man-made systems. For instance, previous reports conclude that protein-protein interaction networks are scale-free (Barabasi and Albert 1999; Albert 2005), small-world (Aloy and Russell 2004) and disassorted (Khor 2010). Furthermore, it was suggested that highly connected proteins in these networks (hubs) are more likely to be essential for cell survival (Albert, Jeong et al. 2000). Networks possess inherent topological properties, which are also transferrable to biological systems, such as functional robustness (Barabasi and Oltvai 2004). Biological networks can be constructed in different ways (Merico, Gfeller et al. 2009). One possibility is to assemble the network de novo from direct experimentally validated interactions. This particular option has been used throughout all network analysis steps in this thesis. Another option is presented by applying known interactions to an –omic dataset, either manually or systematically using pathway-analysis software (e.g., Ingenuity Pathways Analysis), which can be useful for hypothesis-generating experiments. The third possibility is by reverse engineering to generate a subset of networks ab initio which can predict the dynamics of the system, given that sufficient data are compiled across a number of perturbations to enable network modeling (e.g. drug concentrations, enzyme kinetics). This option is often used in quantitative pharmacology and clinical pharmacology studies (Khalil, Brewer et al. 2010). Additionally, networks can include heterogeneous information such as drug-target networks (Yildirim, Goh et al. 2007), where the drug targets are connected only if they are influenced by the same drug or disease networks (Goh, Cusick et al. 2007), where disease genes associated with similar genetic diseases are clustered together. Association of genes with diseases can be achieved using protein interaction networks (Navlakha and Kingsford 2010). Furthermore, the value of PPI networks for the investigation of novel drug targets has been underlined in a number of reviews in recent years (Ruffner, Bauer et al. 2007; Vidal, Cusick et al. 2011).

Network analysis turns into integrated network analysis when network algorithms allow the visualization of different existing interactions, newly discovered relationships, activation and inhibition and superimposition of additional properties beyond the known components (Merico, Gfeller et al. 2009). Nodes or edges can be weighted with qualitative or quantitative information to integrate multiple sources of high-throughput data and connect datasets with seemingly different origins into a refined systems biological dataset ready for in-depth investigation (Sauer, Heinemann et al. 2007). Integrated network analysis then allows the investigation of topological modules, when only the structure in the network is considered (Figure 2a). Further applications are the search for functional modules, consisting of proteins with similar function and similar regulation which change in a similar way after stimulation (Figure 2b). Disease modules of proteins expressed in a similar fashion during the course of disease can also be investigated after gene expression experiments (Figure 2c).

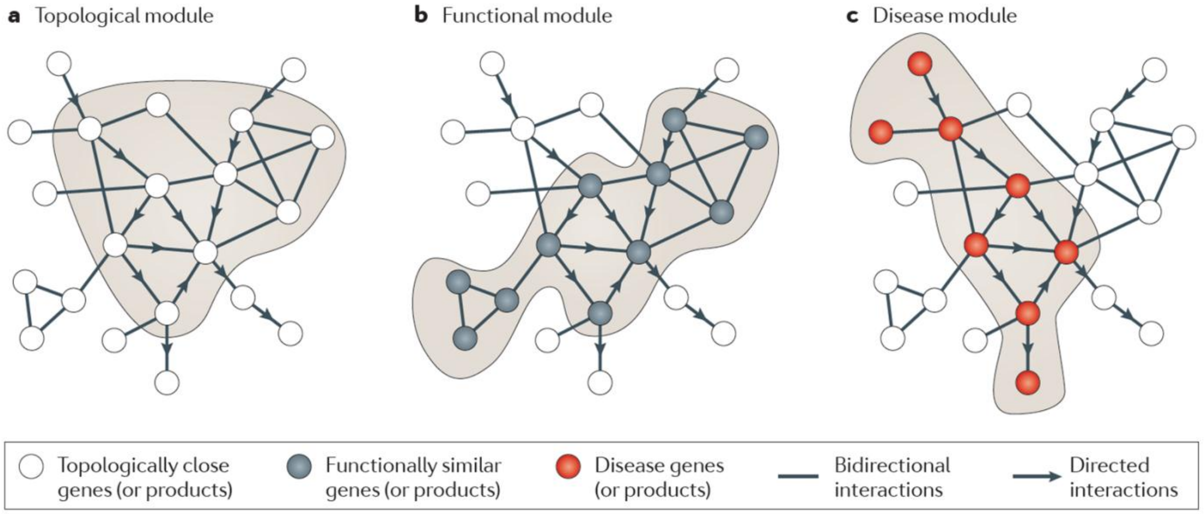

a Topological module
b Functional module
c Disease module
○ Topologically close genes (or products)
● Functionally similar genes (or products)
● Disease genes (or products)
Bidirectional interactions
Directed interactions

Figure 2. Network module types

Topological modules (a) are based purely on the network structure. They correspond to local neighbouring clusters, where the nodes of the module show a higher connectivity within the module when compared to nodes outside the module. Functional modules (b) are comprised of nodes (genes or proteins) with a related function. Functional module rely on the hypothesis that nodes involved in closely related functions tend to interact more often with each other and are therefore located in the same network cluster. A disease module (c) represents nodes the changes of which (mutations, deletions, expression changes) are associated with a disease phenotype, shown here as red nodes. Thus, different characteristics of the nodes lead to different interpretation of the network analysis results. The Figure was extracted from Barabasi et al (Barabasi, Gulbahce et al. 2011)

3.3 Protein-Protein Interaction (PPI) Networks

The most immediate effectors of cellular changes are the regulated activities of proteins (e.g., enzymes, receptors, transcription factors, etc.). Therefore, the analysis of their interactions presents a systematic approach to understanding signaling mechanisms at the crucial final level of cell response: the proteome. Protein-protein interaction networks comprise of nodes consisting of proteins and edges representing physical binding interactions between the proteins (Pieroni, de la Fuente van Bentem et al. 2008). Two proteins are connected by an undirected edge if there is enough experimental support available that they physically bind to each other. Genomic-based studies such as genome-wide profiling for mutations, microarray, RNA sequencing and genome-wide association studies can define genes and loci associated with disease and provide targets for further analysis (Altshuler, Daly et al. 2008; Woollard, Mehta et al. 2011). However, they have limited prospects for creating clinical prognosis or influencing the discovery of new drug targets, as the correlation between DNA, RNA and protein expression is poor (MacKay, Li et al. 2004; Maier, Guell et al. 2009). Due to changes on mRNA level, such as RNA degradation or silencing and other conditions such as RNA structure or ribosome density and occupancy, it is not always possible to predict the results of particular cell stimuli on the whole cell system (Maier, Guell et al. 2009). With evolving new technologies for high-dimensional screens of the proteome it is now possible to detect important changes in thousands of proteins with differential expression during the course of disease. On the other hand, the proteome of particular cells, tissues and organisms has been assembled and improved over time so that the confidence in the measured interactions can also be ascertained.

PPI networks have been assembled for a wide range of organisms, for bacteria such as Escherishia coli (Butland, Peregrin-Alvarez et al. 2005), for the yeast Saccharomyces cereveisiae (Uetz, Giot et al. 2000; Ito, Chiba et al. 2001), the fruitfly Drosophila melanogaster (Giot, Bader et al. 2003; Uetz and Pankratz 2004), worm Caenorhabditis elegans (Li, Armstrong et al. 2004) and human (Persico, Ceol et al. 2005; Rual, Venkatesan et al. 2005; Gandhi, Zhong et al. 2006). PPI networks allow a smooth integration of multiple data types and their coupling with topological graph properties to create more accurate models of cellular interactions. Nonetheless, there are still many present challenges such as the incompleteness of the human interactome both regarding the interactions that occur in a specific context and their regulation.

The large-scale analysis of protein interactions is mainly achieved by using two complementary technologies: yeast two-hybrid (Y2H) and affinity-purification mass-spectrometry (AP-MS). The Y2H method is based on the observation that transcription factor activity can be regained from physically separated activation domainst (AD) and binding domains (BD). Both BD and AD are bound to bait and prey proteins, respectively, and the interaction between the proteins is tested by tracing the expression of a reporter gene activated by the fused transcription factor in an yeast cell (Fields and Song 1989). The reporter gene is transcribed only if the AD and BD domains of the transcription factor come in close proximity after interaction of their fused bait and prey proteins (Figure 3). Large-scale analysis is achieved by mating different yeast strains each expressing a different bait and prey protein (Chien, Bartel et al. 1991).

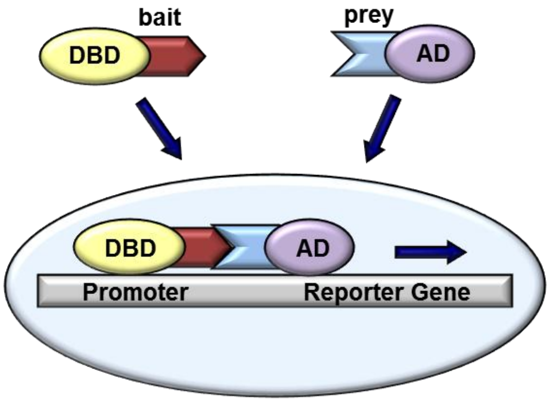


Figure 3. Yeast-two-hybride method

The DNA-binding domain (DBD) of a transcription factor is bound to the bait protein, whereas the activation domain (AD) is bound to the prey-protein. The promoter of the reporter gene is only activated when both DBD and AD bind to it after interaction between the bait and prey proteins.

In contrast to this approach, AP-MS is based on biochemical purification of protein complexes using affinity columns and subsequent identification of complex proteins using mass-spectrometry (Collins and Choudhary 2008) (Figure 4). A bait protein is fused with a tag, which is recognized by an antibody bound on the wall of the affinity column. When the protein is eluted through the column, it binds to the antibody via its tag. All other complex members bind to the initial bait as well and are maintained on the affinity column after a washing step. Then, the proteins can be separated according to their molecular weight using gel electrophoresis and identified with a mass spectrometer. Thus, AP-MS enables the detection of protein complexes under approximately physiological conditions with high sensitivity and precision.

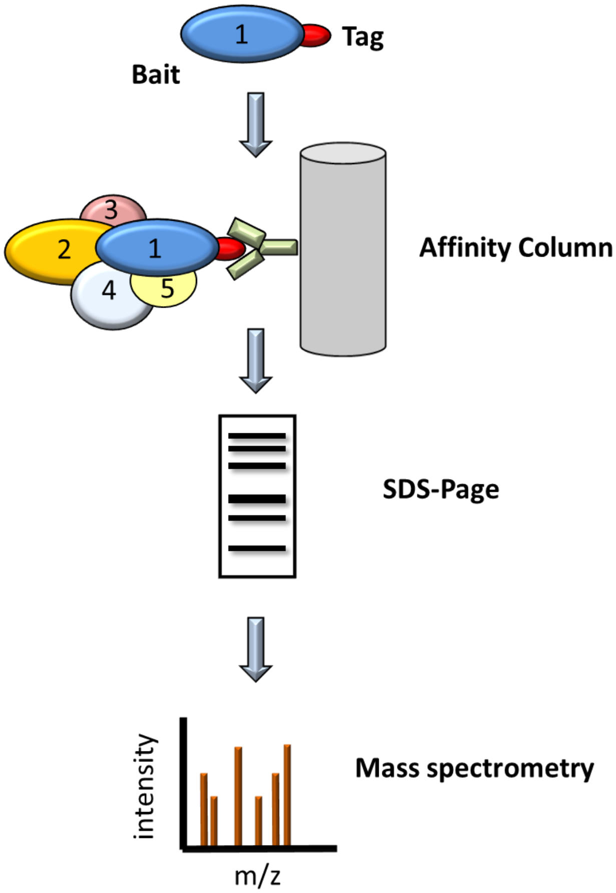


Figure 4. Affinity purification coupled to mass-spectrometry

The AP-MS method comprises of a purified bait protein coupled to an antibody-tag. When the protein runs on an affinity column, the tag binds to its antibody and all proteins involved in the same complex as the bait protein are also immobilized on the column. After a gel electrophoresis step (SDS-Page: sodium dodecyl sulfate polyacrylamide gel electrophoresis), the proteins are separated according to their molecular weight. Subsequently, the protein probe is analyzed using a mass spectrometer, which identifies all complex proteins based on their mass-to-charge (m/z) ratio and intensity spectrum.

While Y2H approach is more adapted to the detection of binary protein interactions, MS-AP has been established mainly for the investigation of protein complexes. Both methods are prone to errors by detecting false positive and false negative interactions. Estimations of the number of interactions occurring in specific cell types have already been proposed. The size of the human interactome has been estimated between 154,000 and 369,000 (Hart, Ramani et al. 2006) or as large as 650,000 interactions (Stumpf, Thorne et al. 2008). Surely, not all interactions are present at all times and many interactions are specific to specific cell types and cell conditions. Several groups have focused on overcoming the challenges of incompleteness in current interactome maps by using various approaches for creating more optimal and objective interaction detection strategies (Lappe and Holm 2004; Schwartz, Yu et al. 2009). Nonetheless, incomplete as the interactome is, it still presents the only way to integrate information on a systems biological level and investigate cell signaling under different conditions.

3.4 Functional modules in biological networks

The availability of high-throughput data such as DNA microarray experiments and proteomics, make it easier to provide large amounts of data for systems biological analysis. Constant development of the proteomics field has led to advances such as quantitative proteomics and phosphoproteomics, but their high costs and complicated analytical procedures are the main reason for ongoing use of qualitative approaches, producing only a list of proteins identified in a sample. The interpretation of these lists is often complicated and data is usually fractionated and lacking some important signaling players due to experimental limitations and outer conditions. Single proteins or subsets of interest are then being selected for further analysis based on predefined pathway data or pre-knowledge from other experiments. A backbone of the interactions between these proteins is crucial for understanding the network context of the analyzed sample. This backbone is supplied by the PPI network of the analyzed organism, which naturally also contains pathway information. Creating a biological network from the data and further exploring functional modules inside this network is a very comprehensive approach for systems biological analysis of transcriptome and proteome data. A functional module is defined as a group of proteins with a similar biological function. Based on this definition, their regulation and expression changes are also similar. Subsequent experimental testing for functional enrichment can provide more insights into the important biological processes overrepresented in the analyzed network. Although a lot of effort has been invested in developing methods for improving the quality of measurements, functional analysis methods have just started to gain attention in the systems biological field. Bioinformatics has helped immensely for developing unbiased data-driven approaches for focusing on small subsets of interesting functionally relevant proteins in large-scale datasets (Ideker, Thorsson et al. 2001; Scott, Perkins et al. 2005; Dittrich, Klau et al. 2008; Beisser, Klau et al. 2010; Zheng and Zhao 2011). Among the algorithms for searching functional modules, there are two distinguishable groups: heuristic approaches and exact approaches (Wu, Zhao et al. 2009). Both approaches use an integrated PPI network as a backbone for their analysis and a weight for nodes and/or edges of the network. While heuristic approaches search for high-scoring subnetworks using a high number of iterations and never reach an optimal solution, exact approaches focus on developing an optimization model and algorithm which then searches for optimal and suboptimal solutions in the pre-defined network. Since the problem of finding high score responsive functional modules from a large interaction network is NP-hard (Ideker, Ozier et al. 2002; Dittrich, Klau et al. 2008) regarding its computational complexity, most computational methods are heuristic approaches. The extracted modules are never optimal and are thus prone to false interpretation, furthermore, the resulting subnetworks can’t be reproduced in the exact same way in a second run. One of the first developed heuristic approaches based on the simulated annealing algorithm was proposed by Ideker et al (Ideker, Thorsson et al. 2001; Ideker, Ozier et al. 2002), followed by a number of heuristic algorithms (Scott, Perkins et al. 2005; Scott, Ideker et al. 2006; Guo, Wang et al. 2007; Liu, Liberzon et al. 2007; Nacu, Critchley-Thorne et al. 2007; Ulitsky and Shamir 2007; Ulitsky and Shamir 2009). By contrast, exact approaches such as mathematical programming based methods proposed in (Dittrich, Klau et al. 2008; Wang 2008; Qiu, Zhang et al. 2009) identify maximally-scoring subgraphs in reasonable time from real molecular networks.

Identification of functional modules can be optimized by adding functional information to the edges in the PPI network. Often during analysis it is important to explore not only the similarity of expression patterns of these proteins, but also their functional similarity, which can be calculated based on Gene Ontology (GO) (Frohlich, Speer et al. 2007). During this thesis the GO term annotations of proteins along with the module detection algorithm proposed by (Dittrich, Klau et al. 2008) will be taken into consideration for the functional module search in various networks.

3.5 Systems biology of cell networks: human platelets

Platelets play a key role in hemostasis and represent a central target for research in many pathophysiological processes, including cardiovascular diseases, inflammation, host immune response and metastasis (Varga-Szabo, Pleines et al. 2008; Leslie 2010). These anucleate blood cells originate from megakaryocytes and have a short life span of about 10 days. They are activated by injury of the vessel wall, which causes platelets to adhere to the injured surface, aggregate and build a firm thrombus with the help of surface-adhesion molecules primarily from the family of integrins (Varga-Szabo, Pleines et al. 2008). By secretion of factors such as thromboxane A2 (TXA2) and ADP more platelets are gathered at the damaged endothelium (Broos, Feys et al. 2011). In a tight balance, platelets control the initial steps of hemostasis and thrombus formation and play a key role in pathological processes such as atherosclerosis. The balance between platelet activation and inhibition ensures the optimal functionality of hemostatic mechanisms. Disturbances of this system are involved in the most common cardiovascular diseases: thrombosis, stroke and myocardial infarction (Furie and Furie 2008). Furthermore, many key platelet receptors are associated with genetic diseases and the investigation of the phenotype of diseases patients helped to identify the role of each receptor in platelet signaling (Lambert 2011). Some examples are the Bernard-Soulier syndrome associated with the vWF-binding receptor GPIBA (Pham and Wang 2007), which is responsible for initial platelet adhesion to the damage endothelial wall and Glanzmann throbasthenia, where patients suffer from severe bleeding periods due to the mutation in the ITGB3 integrin in platelets (George, Caen et al. 1990). Investigation of platelet signaling has been one of the main focus of cardiovascular research for decades but only in recent years advances in covering the platelet proteome and investigating qualitative and quantitative changes in platelet proteins (Senzel, Gnatenko et al. 2009). Platelets have not been the only blood cell line which has been investigated using proteomic approaches. Recently, the red blood cell proteome and interactome was detected (D'Alessandro, Righetti et al. 2010) along with separate proteome analysis of all six blood constituents: plasma, T-cells, monocytes, platelets, neutrophils, erythrocytes (Haudek, Slany et al. 2009). Further analyses include the monocyte proteome (Castagna, Polati et al. 2012) and the T-cell proteome under different conditions (Wollscheid, Watts et al. 2004).

Proteomic analysis is crucial in platelets because genetic manipulations are excluded due to the anucleate nature of these cells. However, they contain a pool of mRNA which can be spliced and translated in a regulated manner (Denis, Tolley et al. 2005; Dittrich, Birschmann et al. 2006; Schwertz, Tolley et al. 2006; Rowley, Oler et al. 2011).

1 satz noch das transcriptome mit proteome verbindet

Protemic analyses of platelet signaling have thrived as an effect of the development of new proteomic technologies for proteomics and phosphoproteomics (Walther and Mann 2010). A number of investigations on platelet subcompartments have already been published (microparticles, alpha granules, membranes and the secretome (Senzel, Gnatenko et al. 2009)). Additional studies have concentrated on the platelet phosphorproteome as an important guideline to changes in platelet signaling (Zahedi, Lewandrowski et al. 2008). However, signaling events have not yet been investigated in a network context. Research studies have been focusing on unraveling platelet signaling but mostly on the level of specific molecules and subparts of pathways rather than the platelet proteome as a whole (Purvis, Chatterjee et al. 2008). With the number of performed studies constantly growing, there is a need for a database combining multiple sources of platelet data. This was achieved during this thesis by the development of a platelet-specific knowledge base called *PlateletWeb*. Integrating information on known platelet studies with human PPI and phosphorylation data, network analysis of platelets is now possible.

3.6 Aim of this work

This thesis is mainly focused on a broad analysis of the platelet proteome and phosphoproteome, platelet signaling events and developing approaches for systems biological analysis of proteomic datasets.

The first topic is the assembly of a platelet-specific platform: *PlateletWeb*. In this database, the human PPI network was established, based on interaction and phosphorylation information from literature. Platelet-specific proteomic and transcriptomics studies were manually collected. Drug and disease information along with information on phosphorylations offer various opportunities for systems biological analysis of platelets. This allowed new insights on platelet activation and signaling leading to a first author publication in Blood (Boyanova, Nilla et al. 2012).

I used the assembled platelet-specific information to analyze platelet signaling in a network context by using integrated network analysis. With the help of kinase predictions for experimentally validated platelet phosphosites, I introduced hypothesis for new drug targets and mechanisms of activation in human platelets. I analyzed different aspects of platelet signaling such as integrin signaling (DOK1) and ADP signaling and systematically investigated the networks obtained from *PlateletWeb*.

The third topic includes the search for optimal functional modules in proteomic data a module detection algorithm enhanced with functional information as network edge values and the curated human interactome as a backbone for integrated network analysis. I performed the validation of this method by investigating small and medium-sized protein networks of various cell types and testing the biological importance of the obtained functional modules for these cells.

Finally, I was involved in transferring the module detection algorithm from qualitative to quantitative phosphoproteomics data. At first tested on platelet phosphoproteomics information, the algorithm was further applied to a site-specific phosphoproteomic dataset from human embryonic stem cells to obtain a kinase-substrate network revealing the regulation of differentiation in these cells.

4 Materials and Methods

4.1 Human proteome, interactome and phosphoproteome assembly

Data from various mass spectrometry studies published in recent years together with a first catalogue of the platelet proteome (Dittrich, Birschmann et al. 2008) were used as sources for a new comprehensive platelet proteome. Studies of unfractionated platelets were included along with studies of specific platelet sub-compartments such as plasma membrane, secretome and microparticles. Furthermore, literature-curated information was extracted from the NCBI GeneRifs (Maglott, Ostell et al. 2007) and filtered for new platelet proteins using the keyword “platelet”. Platelet transcriptome data included a previously performed SAGE (Serial Analysis of Gene Expression) analysis of human platelets (Dittrich, Birschmann et al. 2006). Additionally, information about platelet-specific proteins was extracted from main databases such as Uniprot and HPRD, where tissue-specific information for some proteins is given in a separate Table.

A detailed listing of all platelet data sources used is available in Supplementary Materials (Supplemental Table 1).

Information on human protein-protein interactions (PPI) was retrieved from the Human Proteome Reference Database (HPRD) (version 9.0, 04/2010) (Keshava Prasad, Goel et al. 2009) and the Entrez Gene NCBI server (Maglott, Ostell et al. 2007) (accessed 12/2010). The NCBI server provided interaction information from BioGRID and BIND additionally.

The interaction data was combined with data on protein phosphorylation from HPRD (version 9.0) (Keshava Prasad, Goel et al. 2009) and PhosphoSite (accessed 01/2011) (Hornbeck, Chabra et al. 2004) along with kinase predictions for platelet-specific phosphoproteome data (Zahedi, Lewandrowski et al. 2008) using the NetworKIN algorithm (Linding, Jensen et al. 2007; Miller, Jensen et al. 2008). An interaction network was created using data for interacting proteins retrieved from all these multiple sources. Similarly, a phosphorylation network consisting of all kinases and their substrates (which may also be kinases) has been assembled where the degree of each kinase equals the number of substrates it phosphorylates.

After careful annotation, the complete human PPI network consists of 54,218 simple interactions, 4,406 phosphorylation events and 135 dephosphorylation events between 10,916 human proteins.

4.2 Kinase and phosphatase information

The list of human kinases was extracted from Manning et al (Manning, Whyte et al. 2002) and used for reference and validation of the HPRD phosphorylation data. This study has been acknowledged as a golden standard for human kinase annotations and used as reference in a number of studies (Munoz, Low et al. 2011; Hennig, Mikula et al. 2012; Konig, Nimtz et al. 2012).

The catalogue of human phosphatases was acquired in a multi-step procedure. At first all human phosphatases were obtained from the Human Protein Phosphatases PCR Array (Qiagen; 82 phosphatases) and additionally complemented by protein tyrosine phosphatases from a human genome phosphatase study (Alonso, Sasin et al. 2004) (103 phosphatases). The rest of the available phosphatases were manually extracted from the *PlateletWeb* knowledgebase filtered for proteins with the term "protein phosphatase" in their description. Thus, the total number of human protein phosphatases reached 181, with 39 phosphatases associated with a substrate according to HPRD protein modification data. The platelet phosphatases sum up to 39 with 24 of them containing substrate information.

4.3 Experimentally-validated phosphorylation sites and kinase predictions

Platelet-specific, experimentally validated phosphorylation sites were assembled from a recent mass spectrometry analysis of resting human platelets (533 phosphosites) (Zahedi, Lewandrowski et al. 2008), while the rest of the phosphosites were extracted from experiments described in literature (73,734 phosphosites). These sites were mapped to their positions in the original peptide using a perl script. Kinase information for these sites were extracted using a special bioinformatical algorithm NetworKIN (Miller, Jensen et al. 2008). The NetworKIN algorithm combines two different approaches for phosphorylation predictions - consensus sequence motif search (Miller, Jensen et al. 2008) and protein association networks (a network context of kinase and phosphoproteins which makes up to 60-80% of computational capability to assign in vivo substrate specificity) (Linding, Jensen et al. 2007; Linding, Jensen et al. 2008) - in order to generate a full, realistic and statistically more probable prediction of the involved kinase. Two different scores (a motif score and a context score) are calculated for each algorithm and presented in the final results. Specificity is further enhanced by using information on subcellular compartmentalization, colocalization via anchoring proteins and scaffolds, substrate capture by noncatalytic domains, temporal coexpression and kinase-docking motifs. The algorithm consists of two crucial stages. In the first stage neural networks and position-specific scoring matrices are used for phosphosite assignment to one or more kinases using consensus substrate motifs. In a second stage a probabilistic protein network is extracted from the STRING database, where networks are generated using interaction and pathway databases, literature mining, mRNA expression studies and genomic context (von Mering, Jensen et al. 2005). The nearest member of the relevant kinase family in the thus generated network is identified for each phosphorylation site.

To use NetworKIN all protein sequences were identical starting with the sign “>” and the Protein ID (gene name and ID from NCBI, ID from HPRD with the assigned number for different splicing variants, ID from SwissProt). The position of the phosphorylation sites, obtained with the program were also inserted with the according IDs.

For this study, two different versions of the NetworKIN software were used. The two versions use slightly different motifs for the kinase motif search, which is the reason for discrepancies in some predictions. Nevertheless, by combining the two versions of the algorithms, we were able to gain a full view on the kinase predictions regarding specific sites.

4.4 Drugs and disease information

Drug data were downloaded from DrugBank Version 3.0 (Knox, Law et al. 2011), which includes detailed information on both drugs and drug targets. There are two main types of drug categories, according to their stage of development: approved and experimental. Approved drugs have already been introduced in the market, while experimental drugs are still under development. The database contains 4311 human drugs, which have a human drug target in the *PlateletWeb* knowledge base (approved, 1195; experimental, 3015) and act on 2106 distinct human proteins. There are 950 platelet proteins among these drug targets. Notably, drug-protein interactions are not only physical interactions but may also include indirect functional effects.

Genetic disease information for 701 platelet proteins was extracted from HPRD.

4.5 Pathway information

KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were downloaded from the KEGG database (Release 57.0, January 1, 2011) (Kanehisa 2002). It contains information on signaling, enzymatic reactions and biochemical metabolic transformations. The Advanced Pathway Painter v2.26 was used for the visualization of KEGG pathways in the *PlateletWeb* knowledge base. Enrichment analysis of pathways was performed using Fisher’s exact test comparing the number of platelet proteins in the pathway against the number of all platelet proteins annotated in KEGG pathways.

4.6 Transmembrane domain prediction

Transmembrane domains have been predicted using the TMHMM Server, Version 2.0 (Krogh, Larsson et al. 2001) yielding a total of 5107 transmembrane proteins, of which 1158 are platelet proteins.

4.7 Functional gene annotations for platelet-specific analysis

Gene Ontology (GO) information was extracted from the GO database (Ashburner, Ball et al. 2000) (website accessed December 2010) and used for functional enrichment analysis. There are 4,728 platelet proteins annotated with a GO function, which accounts for a coverage of 94%.

GO Enrichment analysis was performed by the BINGO plug-in v2.44 (Maere, Heymans et al. 2005) of the network analysis software Cytoscape (Shannon, Markiel et al. 2003). For the full GO annotation comparison, all platelet proteins with a GO functional annotation in the network were considered (Biological Process (BP): 3,263; Molecular Function (MF): 3,412; Cellular Component (CC): 3,394 of total 5,025 platelet proteins).

4.8 Module detection algorithm extended with functional information

4.8.1 Gene Ontology (GO) functional enrichment

Gene Ontology information was obtained from the GO database (Ashburner, Ball et al. 2000) (accessed May 2011). This is a hierarchically clustered database which annotates biological terms for each known gene into three main categories: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC). The tree-like structure allows ordering the terms according to their functional specificity with the most specific terms found at the far end of each branch. The GO enrichment analyses in this study were performed with the BINGO plug-in (Maere, Heymans et al. 2005) of the visualisation software Cytoscape (Shannon, Markiel et al. 2003). Statistically significant overrepresented terms were selected according to their p-value using a hypergeometric test (Maere, Heymans et al. 2005). All p-values were adjusted for multiple testing using the Benjamini and Hochberg correction (Benjamini and Yekutieli 2001).

4.8.2 GO semantic similarity

Semantic similarity describes the similarity between two GO ontology terms based on various criteria and assigns a value for each two GO term pairs. In this case, scores for the GO semantic similarity were initially calculated based on the predefined method by Schlicker et al (Schlicker, Domingues et al. 2006) where the probability of the most informative common ancestor (MICA) is used for calculating the score. This algorithm was further extended by Frohlich et al (Frohlich, Speer et al. 2007) to calculate functional similarity between "two genes" (GOSim). The *getGeneSim* function in the GOSim package along with the *funSimAvg* as similarity measure (Schlicker, Domingues et al. 2006) determines the average of best matching GO term similarity for both genes. The semantic measurement was calculated for all the three ontologies on all the interactions which are listed in the interactome. These results were then combined together for each of the interaction into one composite score using the BioNet package in R (Beisser, Klau et al. 2010).

4.8.3 Module detection algorithm

For network analysis we applied a recently devised algorithm (heinz, heaviest induced subgraph), which computes provably optimal and suboptimal solutions to the maximal-scoring subgraph (MSS) problem in reasonable running time using integer linear programming (ILP) (Dittrich, Klau et al. 2008). The algorithm is based on the software dhea (district heating) from Ljubi´c et al. (Ljubi?, Weiskircher et al. 2006). We have extended the C++ code in order to generate suboptimal solutions and have created several Python scripts to control the transformation to a Steiner tree problem, the use of dhea and the retransformation to a PPI subnetwork. The dhea code uses the commercial CPLEX callable library version 9.030 by ILOG, Inc. (Sunnyvale,CA) The analysis of a network obtained by combining data from expression profiling study of lymphoma patients with the comprehensive interactome data from HPRD was performed previously by Dittrich et al (Dittrich, Klau et al. 2008). In this prospect, the p-values are derived from the analysis of differential expression between two tumor subtypes as well from the analysis of survival data by cox regression for each node in the interaction network. The main idea is to identify functional modules in the PPI network, sharing common cellular functions. In order to achieve this, a maximally scoring network is devised along with the scoring of the nodes in the network to be identified.

This algorithm was applied to various proteomic datasets individually to extract functional modules based on previously calculated node and edge scores. When no edge scores are introduced, proteins in the sample are assigned a constant positive node score, while the rest of the interactome proteins obtain a constant negative score.

4.8.4 Score transformation of network edges

To obtain an overview of the functional information in the PPI network and test the proportion of signal in the calculated GO scores (for BP, MF and CC) we compared the GO-similarity measures in the actual PPI with that measured on the background sets created by rewiring the edges in the network. The human PPI interactome was randomized twice, keeping the number of occurrences for each gene constant. We calculated the empirical p-values for each GO score (of each interaction) based on the performed randomization. The obtained p-values were then used for fitting a BUM model to the interactome, which divides the scores into a signal and a noise component. The Π-upper value represents the proportion of signal in the network. In a further step, scores for the module detection algorithm were calculated based on the BUM model. This procedure was performed on the BP, MF and CC semantic similarity scores separately and the obtained algorithm scores were assigned to the interactions and used in further analysis.

4.8.5 Calculation of functional interaction scores

To investigate the network from all aspects simultaneously, a combination of the three main ontology scores was needed. Therefore we used a method of p-value aggregation, previously applied in microarray analysis to combine different sets of data (Dittrich, Klau et al. 2008) . The calculated empirical p-values for all three ontologies were thus combined into a single p-value. The aggregated p-values from the BUM model were converted to network edge scores (= protein scores) using the BioNet package (scoreFunction)(Beisser, Klau et al. 2010). We calculated a threshold p-value (FDR), which controls the false discovery rate for the positively scoring p-values. P-values below this threshold are considered to be significant and will score positively whereas those above the threshold are assumed to have arisen from the null model and will be assigned negative scores. The edges missing all three ontology annotations (2456 edges) were assigned the background distribution of the interactome network (=average of all edge scores).

4.8.6 Generation of network node scores

The network node scores were calculated based on the presence or absence of a protein in the measured proteome sample and the interactome edge scores. All proteins from the sample were assigned a positive value, calculated as follows: **Node score (sample) = -avg (scores of connecting edges),** while the rest of the interactome proteins are given the following value: **Node score (rest of interactome nodes) = - avg (all edge scores).** All node scores of the sample and the rest of the interactome are put together to form a single node score file.

4.8.7 Constraints of the algorithm solution

The module detection algorithm can ensure that all proteins in a sample are found in the resulting network. This is achieved by adding a constant value to the protein score. In the case when the algorithm was used without functional interaction scores, there was no score applied to the edges and all proteins from the sample were given a constant value ("+5" to the H9N2-virus infected cells network and "+10" to the T-cells proteome network) to assure that they all appear in the final solution. The rest of the interactome was assigned a negative value of -1.

4.9 Proteomic studies for identification of functional modules

Data for the module detection analysis was extracted from two main studies, including datasets of different size to optimally test the module identifying algorithm enhanced with GO functional information. One of the datasets was from H9N2 virus infected human gastric cells (Liu, Song et al. 2008), where 22 proteins were identified using mass spectrometry. The second dataset consisted of blood constituents (Haudek, Slany et al. 2009), from which the fraction of T-cells was chosen for the analysis. The study contained 970 T-cell proteins.

4.10 Functional module analysis of quantitative phosphoproteomic data

The used dataset originated from a human embryonic stem cell (hESC) study (Rigbolt, Prokhorova et al. 2011). A total of 6521 proteins were identified in the original dataset using Stable Isotope Labeling by Amino acids in Cell culture (SILAC) method, of which 5765 proteins were mapped in the *PlateletWeb* database. From these, 205 were found to be kinases identified in the study. Site-specific phosphorylation changes were measured at four times points after stimuliation with non-controlled medium (NCM) (30 minutes, 1 hour, 6 hours, 24 hours). Embryonic stem cells were stimulated to differenciate using NCM, which is lacking the needed factors to sustain a pruripotent state. The SILAC ratios were calculated to obtain a value for the differential phosphorylation between the treated and the control cells. The SILAC ratios were then transformed into site-specific node scores and the functional module detecting algorithm was used to identify the time-specific response module of phosphorylation signaling during the hESCs differentiation.

4.11 Statistical analyses

Fisher’s exact test (two-sided for kinase enrichment) was used for all enrichment analyses. All p-values were adjusted for multiple testing families using the Benjamini & Hochberg approach

(Benjamini and Yekutieli 2001). Adjusted P-values lower than 0.05 were considered significant. All statistical analyses were performed with the statistical analysis software R version 2.13.0 (R Development Core Team 2011).

Wilcoxon rank sum test was used for drug enrichment analysis after splitting the drug targets into three separate groups according to the drug type affecting them.

Functional enrichment analysis was performed with the Cytoscape plug-in BINGO. BINGO (Biological Network Gene Ontology Tool) is a plugin for Cytoscape. It helps to determine which Gene Ontology categories are statistically over- or underrepresented in a set of genes by using a hypergeometric test. Gene enrichment analysis was performed by the BINGO plug-in v2.44 of the network analysis software Cytoscape 2.8.

4.12 Network visualization

Throughout the study, the visualization of subnetworks is performed by Cytoscape (version 2.8.2). Cytoscape is an open source platform for complex network analysis and visualization (Shannon, Markiel et al. 2003). It is a java based tool and can be run as standalone software on the local computer.

4.13 *PlateletWeb* database search

The analysis of the platelet database was performed using MySQL – a relational database management system. The data was extracted from multiple sources and saved as tables in the database. Gene Identifiers were used for cross-mapping of tables, therefore they were kept unique. Further programs for analyzing and parsing the data include Notepad++, Microsoft Word and Excel. The script language Perl (Practical Extraction and Reporting Language) was applied for parsing annotations and preparing data for inclusion to the database.

5 Results

5.1 Analysis of the platelet proteome

Mass spectrometry is currently a very powerful protemic tool for large-scale analysis of the cell proteome. During recent years proteome analysis on a cell and tissue scale has been increasingly applied (Kanehisa 2002; Hornbeck, Chabra et al. 2004; Keshava Prasad, Goel et al. 2009). Interaction measured between two proteins under specific conditions such as yeast-two-hybrid or affinity chromatography coupled with mass spectrometry, can be a strong indication that these proteins also interact in vivo. This information can then be used to create a network, where nodes represent the proteins in the interactome and edges stand for the interactions between them. Thus, a whole new perspective of the proteome is gained and methods typically used in statistical network analysis can be applied to investigate biological networks.

Platelets are fitting targets for a protemic approach because they lack a nucleus and they can be obtained in high yield. Furthermore, they are easy to separate from other blood cells. A platform combining the known proteomics and transcriptomics information about platelet protein detection has been missing in the field. Therefore, we introduced a knowledge base (*PlateletWeb*) which integrates a various large-scale datasets yielding a comprehensive catalogue of human platelet proteins and their specific regulation mechanisms (Boyanova, Nilla et al. 2012). For example membrane proteins are difficult to extract because they are usually masked by other peripheral proteins and they are low-abundant. Therefore, a large-scale analysis on membrane proteins complemented the platelet proteome information in the knowledge base. Platelet source data includes references, which enable the evaluation of data quality while the various query modes allow data integration, thus providing insights on multiple options such as pharmacological modulation and network analysis. The integrative view on platelet signaling allows a deeper understanding of the signaling mechanisms in human platelets.

During the development of the *PlateletWeb* database, I was involved in gathering platelet-specific information from proteomic studies, filling and validating the database structure of the database and interpreting biological results in reference to disease and drug target associations. My main focus was the mapping of phosphorylation sites and analyzing the regulatory mechanisms of platelet phosphorylation and dephosphorylation events. Experimental data on platelet phosphorylations obtained from a collaborator mass-spectrometry lab of Sickmann et al (Zahedi, Lewandrowski et al. 2008) was additionally analyzed to obtain kinase predictions for the measured sites, which additionally enhanced the regulatory information available in the *PlateletWeb*. Furthermore, I performed an enrichment analysis of platelet drug targets and disease-associated genes and analyzed in detail targets of approved and experimental drugs.

This research was originally published in Blood. All figures in this chapter have been extracted from the publication by (Boyanova, Nilla et al. 2012).

5.1.1 Data assembly of the platelet proteome and interactome information

Investigation of the platelet proteome was performed based on multiple studies available in literature. Data sources were classified into three main categories: proteome studies (21), transcriptome studies (SAGE (Dittrich, Birschmann et al. 2006)) and database information (Uniprot, Generifs NCBI, Global Proteome Machine Database GPMDB, HPRD). Proteins were analyzed individually depending on the source of platelet information (proteome and transcriptome). Detailed annotation resulted in a set of 5,025 platelet proteins creating a comprehensive and reliable backbone for platelet-specific information. Platelet proteins were detected in multiple fractions depending on the type of study. Most studies included whole platelet lysates or membrane proteome analysis. The platelet secretome was also well covered in proteomic studies (Figure 5). One interesting aspect was to analyze which proteins are identified in most studies. Fibrinogen was found most often, followed by filamin A and actinin. Fibrinogen is a plasma glycoprotein, which was probably measured along with the platelet lysates because of its high abundance. Filamin A is involved in actin filament assembly and is therefore critically involved in platelet shape changes, it also links actin filaments to the cytosolic domain of many membrane glycoproteins in platelets through its C-terminal region (Garcia and Jay 2006). Actinin also plays a role in microfilament assembly and has been indicated to play a potential role in setting integrins to a default low-affinity ligand-binding state in resting platelets and regulating α2bβ3 activation by inside-out signaling in platelets (Tadokoro, Nakazawa et al. 2011).

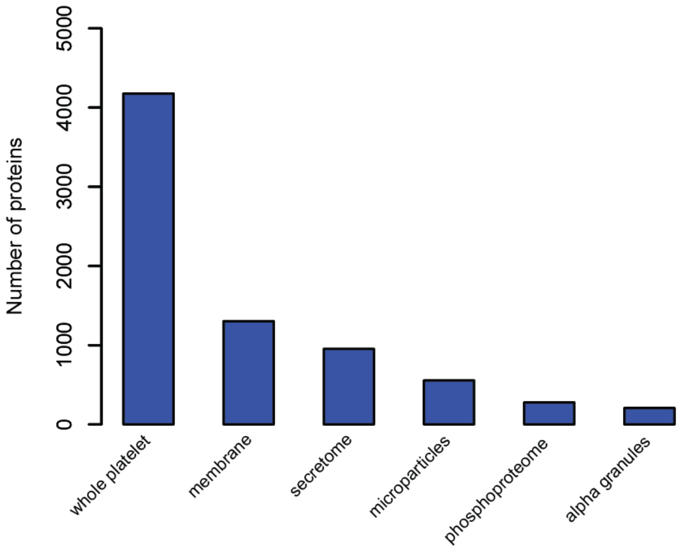


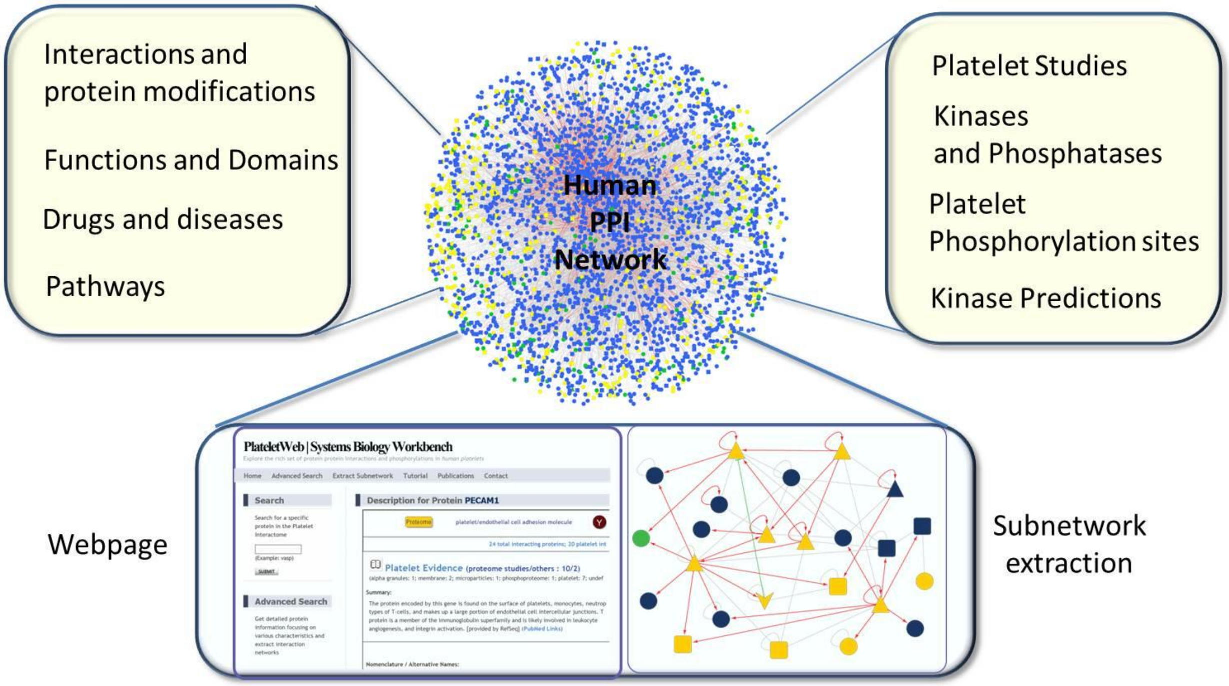
Figure 5. Platelet fractions

Platelet studies were analyzed for the number of proteins extracted in each fraction. Whole platelet analysis was the most abundant fraction, while alpha granules contained the least number of identified proteins. The number of membrane proteins is also quite big, given the difficulty of extracting pure membrane lysates *(Boyanova, Nilla et al. 2012).*

In a previous study (Dittrich, Birschmann et al. 2008), the platelet proteome and transcriptome data was combined with human protein-protein interactions (PPI) to obtain a first overview of the platelet interactome. The PPI information has been extended, by adding new data from manually curated human protein reference database (HPRD), thereby ensuring high quality of annotations. Further information for interactions was added by including phosphorylation events into the *PlateletWeb* database from two major quality databases: HPRD and Phosphosite. The site-specificity and kinase association of each measured phosphorylation site was kept, so that the regulation of the site can also be examined (where known). These phosphorylations are measured in various human cells and information can be extracted in vivo or in vitro, but they are still a very good indicator for existing modifications in human platelets. If a phosphorylation event is known to take place in another cell type and both the kinase and protein are present in platelets, then there is a high chance that this modification also occurs in platelets. To increase platelet-specificity, a platelet phosphoproteome study by Zahedi et al (Zahedi, Lewandrowski et al. 2008) was included to the database, where phosphorylation sites were measured with mass spectrometry in platelets. These 533 phosphorylation sites were especially important to our analysis, as they were measured under basal conditions in resting platelets. Mapping of these sites to the corresponding proteins to identify the exact position represents an important part of my work during the master thesis. During this thesis, I included these phosphorylation sites and their kinase predictions into a network context, made freely accessible with the *PlateletWeb* knowledge base.

To allow systems biological analysis of the newly gathered interactome and phosphoproteome information, information of available kinases and phosphatases was crucial for understanding the regulation of phosphorylation sites in a network context. Acting kinases were assembled from HPRD and Phosphosite but to achieve a comprehensive analysis, more kinase information was included from studies of the human kinome. Manning et al (Manning, Whyte et al. 2002) first introduced a standard list of human kinases in 2002 and proteins in the database were defined as kinases if they were available in this study. Phosphatase data was also added to the knowledge base as dephosphorylation is also a well-established mechanism in cell signaling, triggering either activation or inhibition of the dephosphorylated protein. Again, a reliable source for phosphatases was considered from the Human Protein Phosphatases PCR Array (Quiagen; 82 phosphatases) and the assembly of protein tyrosine phosphatases in the human genome (Alonso, Sasin et al. 2004) (103 phosphatases). Further manual curation of phosphatases was performed using a search in the protein summary for the term “protein phosphatases” yielding a total of 191 phosphatases. The full list of platelet phosphatases is available in Supplemental Table 2.

To allow investigation of pharmacological modulation options, drug information was retrieved from the drug repository DrugBank. Each platelet and non-platelet protein is now available with its targeting drug. Another important aspect of platelet proteins is whether they are associated with genetic diseases therefore data from HPRD on genetic diseases was also included in the database. Functional information added from the Gene Ontology database contributed to the wealth of data available on a single-protein level. Further valuable information included protein domains and transmembrane predictions. To ensure a comprehensive signaling analysis of platelet function, KEGG pathways are also introduced in the database and all contained platelet proteins are additionally highlighted in the KEGG visualization. The integrative information sources can be seen in Figure 6.


Interactions and protein modifications
Functions and Domains
Drugs and diseases
Pathways
Human PPI Network
Platelet Studies
Kinases and Phosphatases
Platelet Phosphorylation sites
Kinase Predictions
Webpage
PlateletWeb | Systems Biology Workbench
Explore the rich set of protein protein interactions and phosphorylations in human platelets
Home
Advanced Search
Extract Subnetwork
Tutorial
Publications
Contact
Search
Description for Protein PECAM1
Proteome
platelet/endothelial cell adhesion molecule
Y
24 total interacting proteins; 20 platelet int
(Example: vasp)
SUBMIT
Platelet Evidence (proteome studies/others : 10/2)
(alpha granules: 1; membrane: 2; microparticles: 1; phosphoproteome: 1; platelet: 7; undef
Summary:
Advanced Search
Nomenclature / Alternative Names:
Subnetwork extraction

Search for a specific protein in the Platelet Interactome

Get detailed protein information focusing on various characteristics and extract interaction networks

The protein encoded by this gene is found on the surface of platelets, monocytes, neutrop types of T-cells, and makes up a large portion of endothelial cell intercellular junctions. T protein is a member of the immunoglobulin superfamily and is likely involved in leukocyte angiogenesis, and integrin activation. [provided by RefSeq] (PubMed Links)

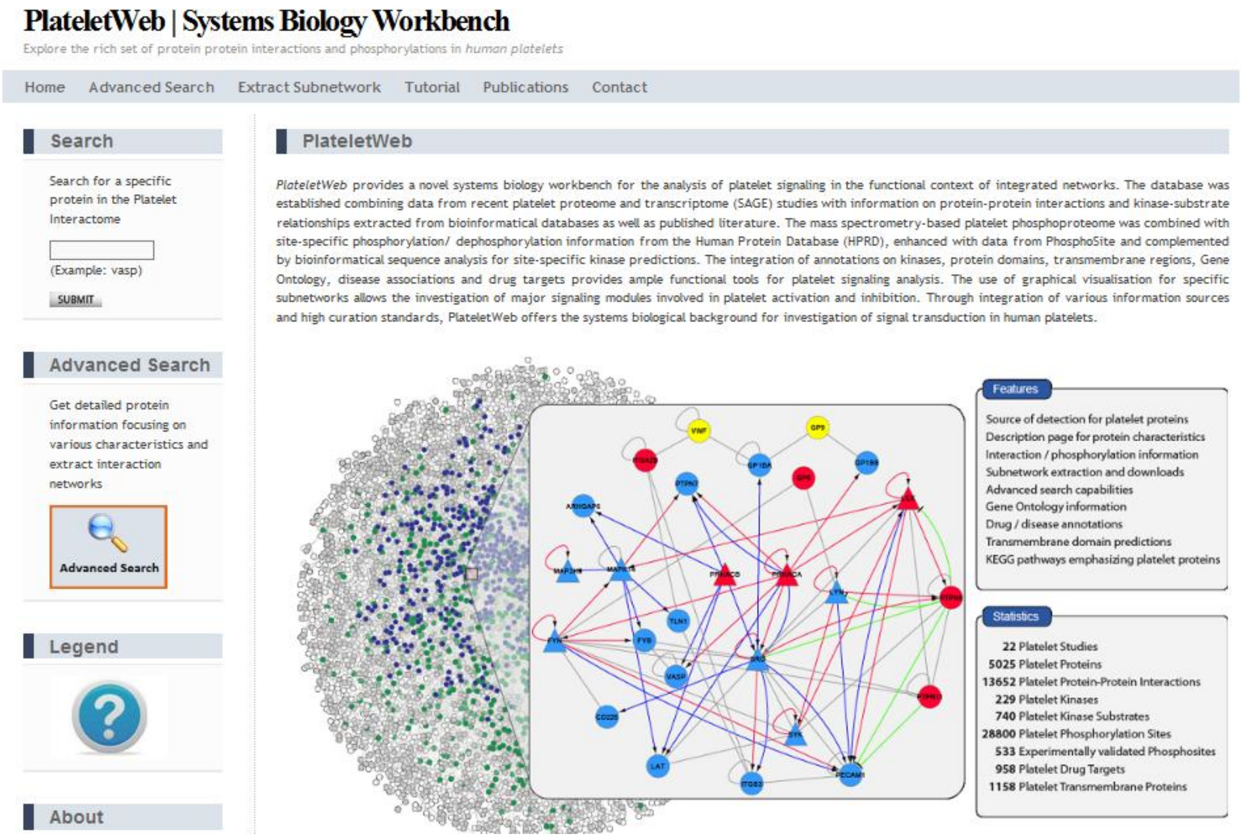
Figure 6. *PlateletWeb* database sources

Integrated information on interactions and protein modifications (phosphorylations and dephosphorylations), functional data, domains, drug and disease association and pathways was added to platelet-specific proteomic data extracted from multiple proteome studies and complemented by kinase and phosphatase information, site-specific phosphorylation sites and kinase predictions for experimentally validated platelet phosphosites. Thus, a comprehensive backbone for the platelet systems biological investigation is now available.

All data is available online on a website (*http://PlateletWeb.bioapps.biozentrum.uni-wuerzburg.de*) with a broad and intuitive interface for functional network analysis including advanced data mining capabilities and the visualization of subnetworks with integrated information on phosphorylations and interactions.

5.1.2 *PlateletWeb* knowledgebase

The *PlateletWeb* knowledgebase was created by integrating multiple sources of data, which allows a first systematic overview of the platelet proteome network with its phosphorylation events and drug regulation. Using the platform it is possible to extract information on a single protein level as well as network context level. A query search for a protein of interest yields results about its interaction partners and physical characteristic. Each interaction is supported with evidence from literature allowing the user to trace back the original source of the interaction. Furthermore, characteristic features from the functional and network context of the protein along with the technique for its identification in platelets (level of detection) are available, representing also the number of studies the protein has been identified in and the fraction from which it was isolated. Information on physical properties such as isoform-specific sequence information, transmembrane domains, isoelectric point and molecular weight are additionally provided. Advanced search for any of these characteristics allows the user to look for a protein of a particular isoelectric point range or a particular weight, making this tool especially useful in the analysis of Western Blot and 2D Gel experiments, where the weight, size and isoelectric point are crucial for identifying proteins from the experiment. Special focus is given to site-specific phosphorylation and dephosphorylation sites, with kinase/phosphatase data if it is available for the given site. Thus, the phosphorylation status of all proteins can be investigated, distinguishing between phosphorylations derived from published literature and those directly measured in human platelets. Additional information on the source study for these modifications enables the user to verify the reliability of the platelet proteomic sources. Advanced search for particular features of interest can also be performed, including combinational search for proteins with specific phosphorylated residues detected on proteome or transcriptome level and containing particular domain information along with functional annotation for a particular GO term. Drug information provides additional insights about platelet proteins associated with specific drugs. As an example, results on pharmacological modification by inhibiting prostacyclin receptors retrieve analogues of prostacyclin (Epoprostenol, Iloprost, Treprostinil). A key-word search in the description of proteins can be used to find associations with diseases. The knowledge base thus allows a comprehensive analysis of the platelet on a single protein level as well as on the scale of network regulation and functional association of signaling components. The main page of the *PlateletWeb* website is represented in Figure 7.



PlateletWeb | Systems Biology Workbench

Explore the rich set of protein protein interactions and phosphorylations in *human platelets*

Search

Search for a specific protein in the Platelet Interactome

(Example: vasp)

Advanced Search

Get detailed protein information focusing on various characteristics and extract interaction networks

Legend

About

PlateletWeb

PlateletWeb provides a novel systems biology workbench for the analysis of platelet signaling in the functional context of integrated networks. The database was established combining data from recent platelet proteome and transcriptome (SAGE) studies with information on protein-protein interactions and kinase-substrate relationships extracted from bioinformatical databases as well as published literature. The mass spectrometry-based platelet phosphoproteome was combined with site-specific phosphorylation/ dephosphorylation information from the Human Protein Database (HPRD), enhanced with data from PhosphoSite and complemented by bioinformatical sequence analysis for site-specific kinase predictions. The integration of annotations on kinases, protein domains, transmembrane regions, Gene Ontology, disease associations and drug targets provides ample functional tools for platelet signaling analysis. The use of graphical visualisation for specific subnetworks allows the investigation of major signaling modules involved in platelet activation and inhibition. Through integration of various information sources and high curation standards, PlateletWeb offers the systems biological background for investigation of signal transduction in human platelets.

Features

Source of detection for platelet proteins
Description page for protein characteristics
Interaction / phosphorylation information
Subnetwork extraction and downloads
Advanced search capabilities
Gene Ontology information
Drug / disease annotations
Transmembrane domain predictions
KEGG pathways emphasizing platelet proteins

Statistics

22	Platelet Studies
5025	Platelet Proteins
13652	Platelet Protein-Protein Interactions
229	Platelet Kinases
740	Platelet Kinase Substrates
28800	Platelet Phosphorylation Sites
533	Experimentally validated Phosphosites
958	Platelet Drug Targets
1158	Platelet Transmembrane Proteins

Figure 7. Screenshot of the main page of the *PlateletWeb* knowledge base

The screenshot represents the title page of the *PlateletWeb* knowledge base, with available information on the basic concept of the website and links to advanced search. The option for extracting subnetworks of a given list of proteins is also available from the main page. Additionally, links to the tutorial, publications and contact are readily accessible along with an available legend, explaining the color coding of various visualizations in *PlateletWeb*. The main network presented, shows the vWF signaling pathway as an example for an integrated network, with all available phosphorylations, dephosphorylations, interactions and the the phosphorylation state of the proteins (blue for platelet proteins phosphorylated in the platelet, red for platelet protein phosphorylated in human cells and yellow for unphosphorylated platelet proteins). Beside the network information, there are also statistics on the content of the database.

5.1.3 Analysis of the platelet phosphoproteome

The first step in analyzing the platelet phosphoproteome was to determine the distribution of serine, threonine and tyrosine phosphorylation sites of platelet proteins, either from literature documentations in other human cells or from the experimentally validated platelet phosphosites. The residue-specific distribution showed a clear majority of serine phosphorylated sites in both human (57.2%) and platelet sites (82.7%), but the amount of tyrosine phosphorylations measured in platelets were lower than the overall distribution (4% when compared to 21.3%). A possible explanation for this has been proposed by (Olsen, Blagoev et al. 2006) where it is mentioned that tyrosine phosphorylations are often found in low abundant proteins. Additionally, their lower stability in phosphoamino acid analysis makes them harder to detect. The amount of serine phosphorylations in the experimental set is a lot higher than in the overall distribution (Figure 8).

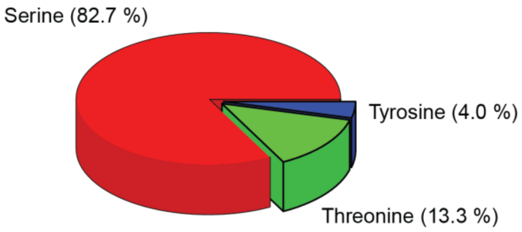


Figure 8. Distribution of experimentally validated phosphorylation sites in platelet proteins

Distribution of serine, threonine and tyrosine phosphorylation sites among platelet proteins. The percentage of each phosphorylation fraction is represented in brackets *(Boyanova, Nilla et al. 2012).*

Phosphorylations can only be fully understood and investigated in a network if information on the phosphorylating kinase is available. Thus, we first analyzed the number of phosphorylated platelet proteins associated with a kinase, representing only 23% (814) of all phosphorylated platelet proteins (3,532). For investigating proteins in the network context, it was essential to extract also data about the regulation of their phosphorylation sites. On the scale of phosphosites, kinase and phosphatase information was available for a total of 3,080 sites. When limiting the study only to phosphosites measured in platelets, information was available for only 69 phosphosites. Therefore, we used a novel network-based algorithm for prediction of potential kinases responsible for the phosphorylation of these sites (Linding, Jensen et al. 2007; Miller, Jensen et al. 2008), which yielded kinase predictions for further 436 sites. The total kinase annotation for experimentally validated phosphosites was thereby extended to 505 sites associated with a kinase (94.5%). When the whole platelet phosphoproteome was taken into consideration, these predictions contributed to 16% of all modifications with available kinase or phosphatase information (Figure 9).

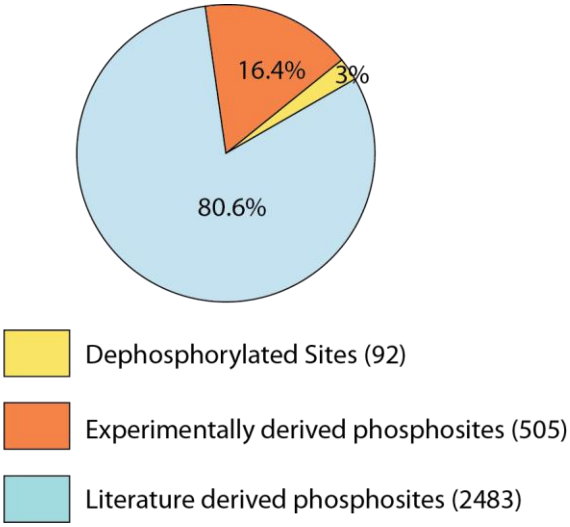


Figure 9. Distribution of modificated sites associated with a kinase/phosphatase.

The number of modificated sites is presented as a fraction of all modificated sites with known kinase or phosphatase (3080). The dephosphorylated sites with known regulation represent the smallest fraction. Kinase predictions for experimentally validated platelet phosphosies account for 16,4% of all sites associated with a kinase, thereby increasing the amount of sites with known regulation *(Boyanova, Nilla et al. 2012).*

From the overall 526 kinases 229 were found in platelets (43,5%). Not all kinases are supplied with substrate information, only about 162 (70.7%) have well-described substrates in the platelet proteome but nearly all (216, 94%) have documented phosphorylation sites. On the side of phosphatases, 73 (38.2 % of 191 total human phosphatases) phosphatases are identified in platelets and 24 of them have characterized substrates. Interestingly, when comparing the number of available platelet kinases and phosphatases, kinases build a clearly larger group, indicating that dephosphorylation is achieved by fewer enzymes, which have a broad range of activity.

Focusing further on the kinase predictions for experimentally validated phosphosites in platelets, we predicted 96 kinases based on the kinaseprediction algorithm (see methods) and 69 of them were identified in platelets. The top 8 kinases with most platelet targets are all detected in the platelet and depicted in Table 1. CDK2 was predicted to phosphorylate 119 distinct proteins on 183 unique sites. The ubiquitously acting casein kinase was listed also among the top predicted kinases with the most substrates. The high number of MAP kinases can be explained by the specificity of the algorithm itself and its bias towards phosphorylation motifs in MAP kinase substrates.

Gene ID	***Gene Symbol***	***Substrates***	***Phosphosites***
1017	CDK2	119	183
5599	MAPK8	91	133
1459	CSNK2A2	87	143
1457	CSNK2A1	87	143
2932	GSK3B	75	102
1432	MAPK14	61	77
5600	MAPK11	60	79
5603	MAPK13	59	75

Table 1. The top 8 kinases predicted by the NetworKIN algorithm

The kinase predicted most often is CDK2, followed by the MAP kinase MAPK8 and the broad specificity casein kinases. MAP kinases are the predominant kinases predicted by the Networkin algorithm.

Protein kinase A (PKA), a kinase which plays an important role for maintaining platelet balance by inhibiting platelets in their resting state, was also predicted to phosphorylate a number of sites. There are 20 proteins with predictions for the alpha subunit PRKACA and 30 proteins with predictions for the beta subunit PRKACB. On the level of phosphorylation sites, there are 40 sites with predictions for either subunit and only 3 of them overlap with already known regulation by PKA from literature (either extracted from HPRD or Phosphosite). For the remaining 37 sites, there is information available only from the predicting algorithm. For instance, the two sites belong to the proteins cyclin Y and syntaxin binding protein 5 (tomosyn). Tomosyn is one of the many components of the neurotransmitter release machinery and binds to syntaxin (Fujita, Shirataki et al. 1998). Syntaxin belongs to the group of t-SNARE proteins. In platelets SNARE proteins mediate the membrane fusion events required for granule cargo release (Graham, Ren et al. 2009). Therefore, tomosyn might play a role in endocytosis of platelets needed after activation and the granule release of transmitters such as ADP. Phosphosites potentially regulated by PKA may be involved in in inhibiting the function of tomosyn while platelets are inactive.

5.1.4 Platelet transmembrane domain proteins

Membrane proteins perform key roles in cell-cell signaling and facilitate the initial steps in cell signaling activation. It is estimated that more than half of all proteins interact either directly or indirectly with cellular membranes (Almen, Nordstrom et al. 2009; Zhang, Naslavsky et al. 2012). From a clinical perspective approximately 70% of all known drug targets are transmembrane plasma membrane proteins (Hopkins and Groom 2002) and therefore predictions for transmembrane (TM) domains were performed for all human proteins in the database. Results from predictions indicated that 23% of the platelet proteome (1158 proteins) have a transmembrane domain, which is in accordance with the number of all human transmembrane proteins constituting 26% of the human interactome. These results correlate well with the study by Almen et al, where 27% of the human genome was estimated to code for transmembrane proteins (Almen, Nordstrom et al. 2009). The next step was to test whether there is an enrichment of a particular group of transmembrane proteins for drug targets according to the number of transmembrane domains they contain.

Comparisons between the number of transmembrane domains of platelet proteins and TM domains of non-platelet proteins showed an underrepresentation of seven transmembrane domain receptors and four transmembrane domain receptors in platelet proteins. A closer overview of these receptors in non-platelet proteins revealed that they are mainly part of neuronal pathways and are therefore missing in platelet cells (Figure 10).

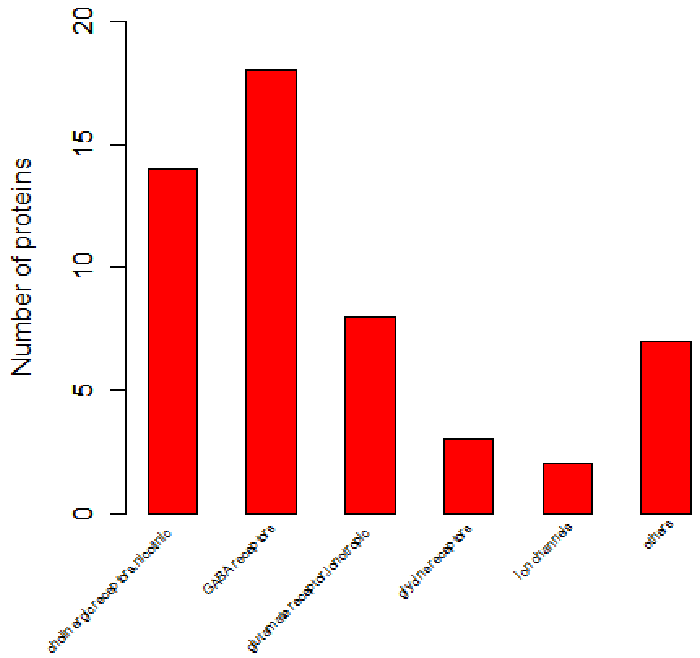


Figure 10. The top functional categories of proteins containing four transmembrane domains

Proteins were categorized into functional classes, which they belong to, and visualized according to the number of proteins with four transmembrane domains contained in each functional group. Most of the functional categories are neuronal receptors missing in platelet, which explains their underrepresentation.

In a second approach the number of identified platelet proteins in membrane fractions was compared with the number of platelet proteins with predicted transmembrane domains. In total, there were 1304 proteins identified in the membrane fraction with membrane proteomics. Interestingly, transmembrane prediction yielded further 532 platelet proteins which were not yet identified in a membrane-specific mass spectrometry study. Out of these, 303 were found on the proteome level with other mass spectrometry studies and among them were many receptors and proteins associated with membrane such as ATPases, cadherins, EPH receptors and integrins. Filtering those proteins for the word "membrane" in their description, yielded 10 platelet proteins identified on the proteome level (Table 2). These proteins consist of channels or outer mitochondrial-membrane proteins associated with transport. Thus, analysis of transmembrane domains using *PlateletWeb* gives indications about yet unidentified potential membrane proteins.

Gene ID	***Gene Symbol***	***Description***	***Number of predicted TM domains***
117532	TMC2	transmembrane channel-like 2	9
7108	TM7SF2	transmembrane 7 superfamily member 2	7
2206	MS4A2	membrane-spanning 4-domains, subfamily A, member 2 (Fc fragment of IgE, high affinity I, receptor for; beta polypeptide)	4
10067	SCAMP3	secretory carrier membrane protein 3	4
10507	SEMA4D	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4D	2
93380	MMGT1	membrane magnesium transporter 1	2
80863	PRRT1	proline-rich transmembrane protein 1	2
8720	MBTPS1	membrane-bound transcription factor peptidase, site 1	1
9804	TOMM20	translocase of outer mitochondrial membrane 20 homolog (yeast)	1
51024	FIS1	fission 1 (mitochondrial outer membrane) homolog (S. cerevisiae)	1

Table 2. Proteins with predicted transmembrane domains, which were not identified in membrane studies

The top 10 platelet proteins with the highest number of predicted TM domains are represented with their gene id, gene symbol and description and their number of predicted domains they contain.

5.1.5 Pathway enrichment in platelets

To characterize a cell type, one has to consider also the various pathways, which most likely play a role in its signaling. After performing a Gene Ontology enrichment analysis on all platelet proteins (Boyanova, Nilla et al. 2012), we tested whether some of the known KEGG pathways contain an increased number of platelet proteins and whether it is significantly different than expected by chance. Pathway information was downloaded from KEGG and platelet proteins were mapped to the KEGG database. A Fisher test was used to determine whether there is a significant enrichment of particular pathways in the platelet proteome by comparing the number of platelet proteins in a given pathway to the number of all platelet proteins found in all pathways. Thus, a higher abundance of platelet proteins would indicate an involvement of this pathway in platelet signaling. Figure 11 represents the top 25 pathways identified and presented according to the significance of their p-values. The most enriched pathway was “endocytosis” (p-value = 2.35×10^{-14}), followed by “regulation of cytoskeleton” (p-value = 1.82×10^{-12}) and “Fc gamma receptor-signaling” (p-value = 3.28×10^{-12}). Further pathways connected to endocytosis are the “SNARE interactions in vesicular process”. Integrin signaling as part of the focal adhesion pathway was also overrepresented, indicating its importance in platelet activation. Pathways associated with inflammation (“Leukocyte transendothelial migration” and “Chemokine signaling pathway”) were enriched for platelet proteins as well. Interestingly, many of the pathways associated with infection and pathogenic entrance were also found in the study: “pathogenic E.coli infection”, “Epithelial cell signaling in Helicobacter pylori infection” and “Shigellosis” as well as disease-facilitating pathways such as “Alzheimer’s disease” and “Parkinson’s disease”.

On the other hand, signaling pathways specific to other tissues such as “Neuroactive ligand-receptor interaction” (P-value = 2.06×10^{-15}) and “Olfactory transduction” were strongly underrepresented in the platelet proteome (Supplemental Table 3).

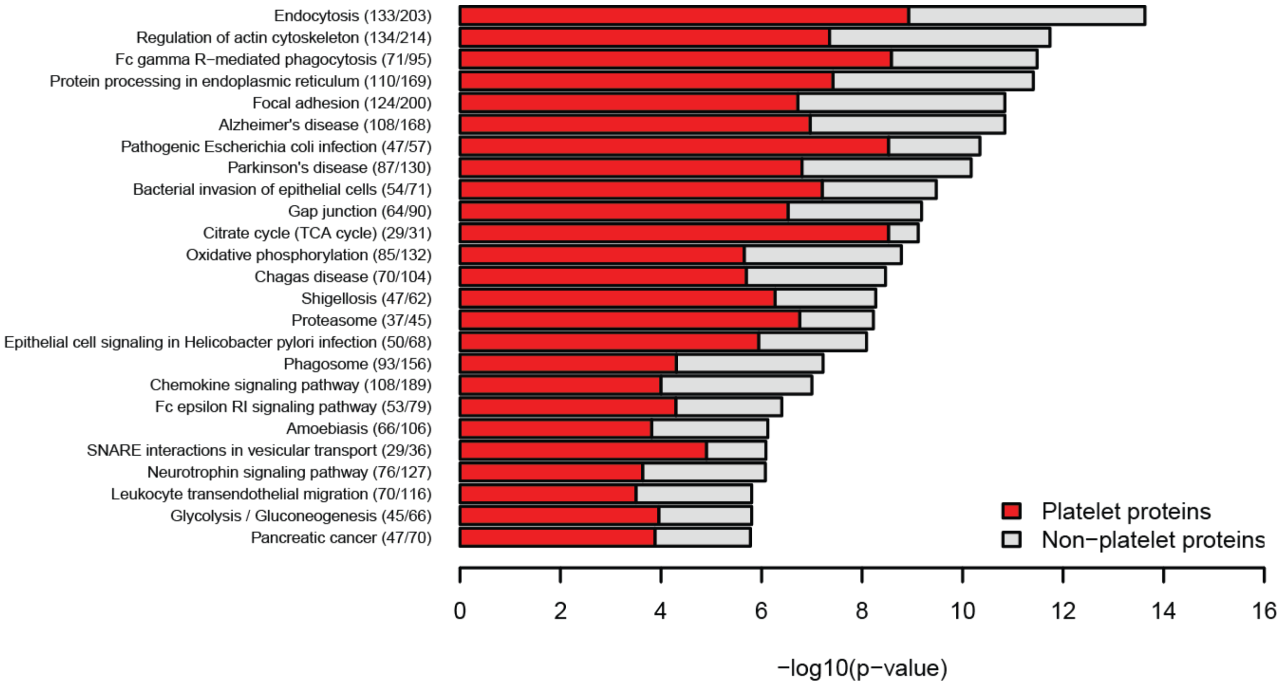


Figure 11. KEGG enrichment of platelet proteins

The top 25 enriched pathways are represented on the x-axis showing the –log10 value of the obtained p-values from the Fisher-test. Endocytosis and regulation of cytoskeleton are highly enriched in human platelets, followed by a number of pathways associated with phagocytosis and disease processes *(Boyanova, Nilla et al. 2012).*

5.1.6 Drug target and disease gene enrichment in platelets

A comprehensive systems biological analysis includes investigation of the pharmacological modulation of proteins. We therefore created an extensive platelet drug target network using data on human drugs and drug targets from the DrugBank database *(Knox, Law et al. 2011)*, including both approved (1195) and experimental (3015) drugs. Approved drugs are FDA-approved pharmaceuticals, available on the market and used in patient therapy, while experimental drugs are predominantly chemical molecules in development which still haven't reached clinical application. The extracted drug dataset consists of 4311 human drugs, half of which (2706) act on platelet proteins. On the side of human proteins, there are 2106 human proteins associated with drugs, from which 950 are platelet drug targets (19% of all platelet proteins). The general statistic showed that 23 % of platelet drug targets are targeted by both approved and experimental drugs, but there is also a large group of targets affected by each drug type individually (Figure 12). Detailed analyses of these drug targets in relation to their topological (platelet interactome) and chemical properties (kinases) may elucidate how well platelet proteins are pharmacologically modulated and what tendencies are followed by the current drug development studies.

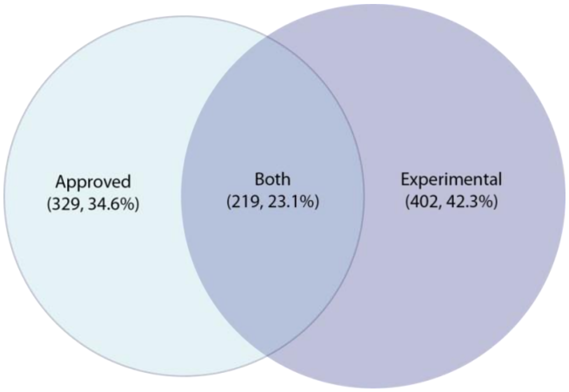


Figure 12. Numbers of platelet drug targets affected by different drug types

The figure represents the proportion of drug targets affected by the two types of drugs: approved and experimental. The two types affect 23% of all drug targets, while the rest is targeted by only one drug type (Boyanova, Nilla et al. 2012).

Topological exploration of platelet drug targets included the investigation of their connectivity in the network. We grouped platelet proteins in the interactome by degree of interaction into non-overlapping intervals (from less than 5 to above 40 interaction partners) and calculated the fraction of drug targets in each group.

The resulting distribution showed an overall increase of drug targets among highly connected proteins in the network (Figure 13A). A Wilcoxon rang test revealed differences in the mean of each group (drug target vs. non drug targets). The number of interactors of drug affected proteins (mean = 11.6) were significantly higher than those not associated with drugs (mean = 6.8; p-value = 1.84 × 10^{-10}, Figure 13B).

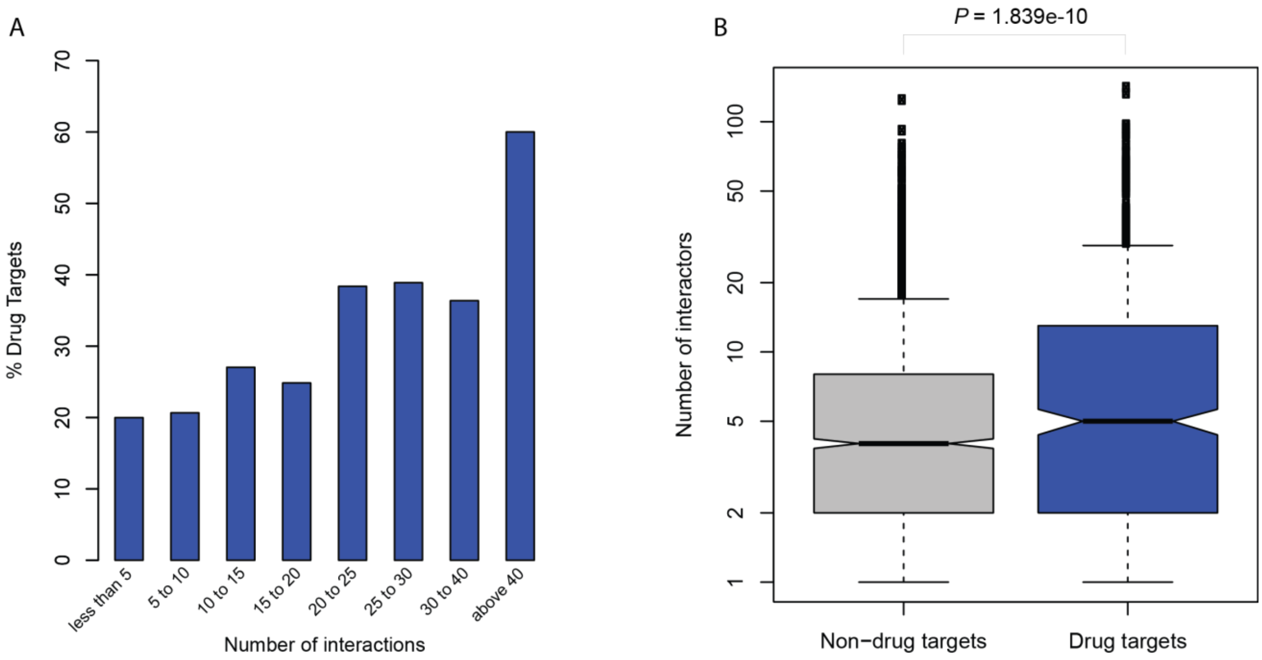


Figure 13. Dependancy of platelet drug targets from their degree

(A) Topological network analysis of drug targets reveals an increase in the number of drug targets among highly-connected platelet proteins. (B) Drug targets and non-drug targets are presented in box-plots according to their number of interactors. The median is higher in the group of drug targets, indicating enrichment of highly connected proteins in drug targets (Boyanova, Nilla et al. 2012).

To investigate the effect of each drug type to the observed enrichment among hub proteins, all drug targeted proteins were separated into three groups: proteins affected by experimental drugs, proteins targeted by approved drugs and proteins affected by both drug types. All three groups of proteins were consequently compared to the proteins without any drug association and a Wilcoxon rank test was applied to evaluate significant differences between these groups. Experimental drug targets were found significantly more often among well connected proteins (mean = 13.1; p-value = 5.79 × 10^{-12}, Figure 14). This tendency decreased when drug targets affected by both drug types were tested (mean = 13.3; p-value = 0.0016). No significant difference could be detected for proteins targeted by approved drugs exclusively (mean = 8.0; p-value = 0.1575).

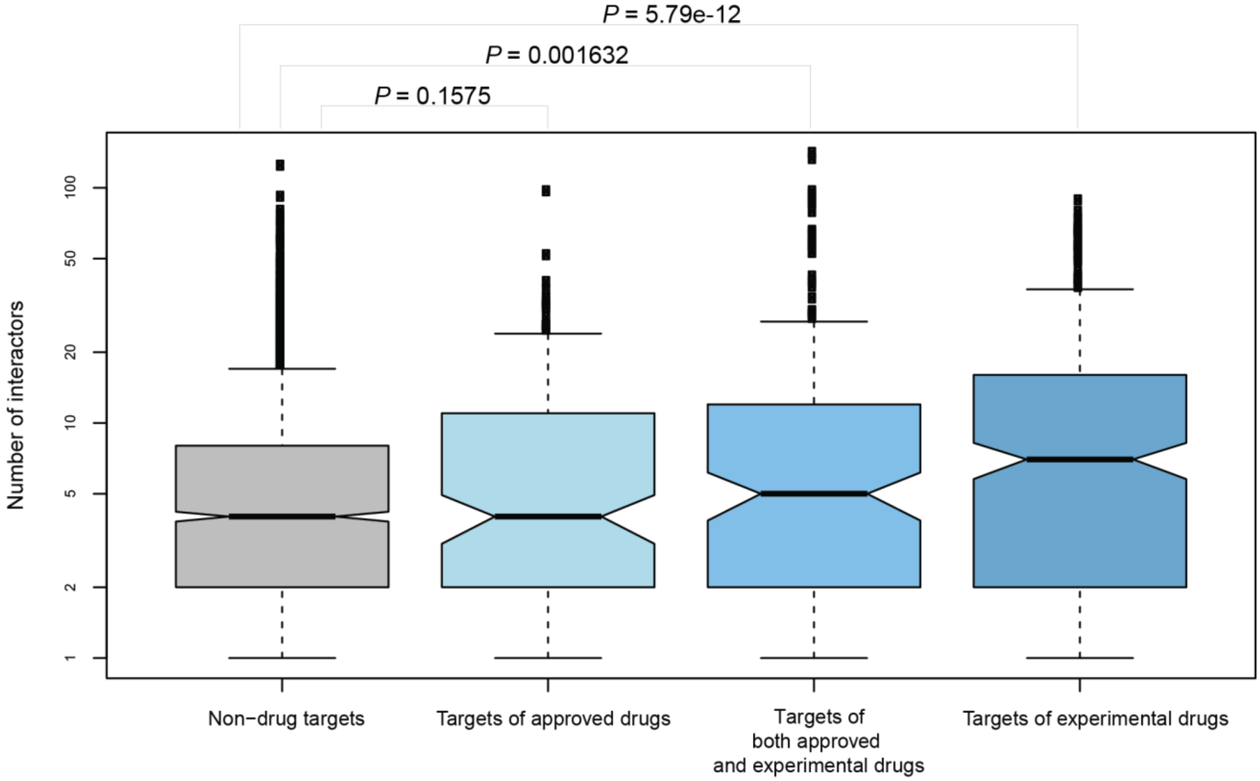


Figure 14. Boxplots of platelet drug targets according to the type of drug

Analysis of drug targets distinguished by the type of targeting drug. Proteins associated exclusively with experimental drugs are more likely to have many interactors when compared to targets of approved drugs or both approved and experimental drugs (Boyanova, Nilla et al. 2012).

These results suggest that experimental drugs under development are more often influencing highly connected proteins in the platelet interactome. For further clarification, we performed a functional enrichment analysis on experimental drug targets. There was an overrepresentation of the process “phosphorylation” and “phosphorus metabolic process” among biological processes (Figure 15A) and “kinase activity” was the top enriched term in molecular functions (Figure 15B). This indicates that proteins targeted only by experimental drugs are in most cases kinases. We suggested that the observed effect of hubs associated with developmental drugs is predominantly due to targeted kinases in the platelet interactome.

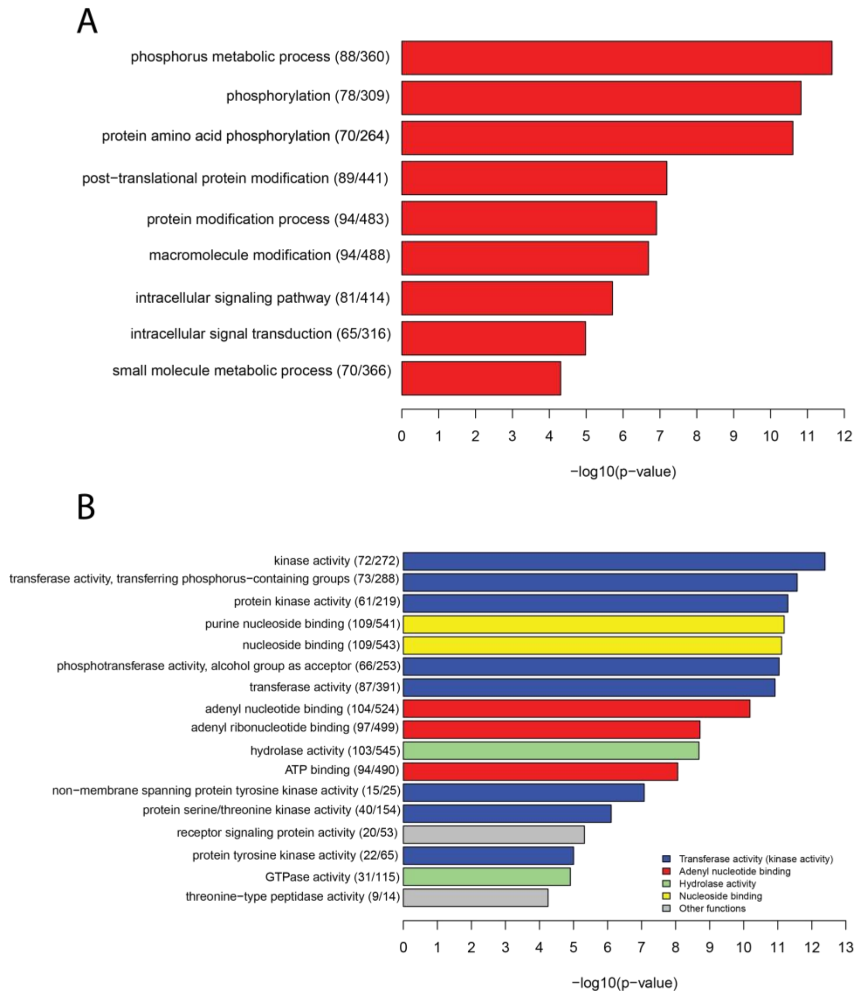


Figure 15. GO enrichment analysis of proteins affected by experimental drugs

(A) The top significant biological process terms and (B) molecular functions of proteins only targeted by experimental drugs are presented according to the –log (p-value) on the x-axis. The terms are grouped according to the parent Gene Ontology term. Phosphorylation processes and kinase activity are enriched. Overrepresented cellular components are not shown (mitochondrion, proteasome core complex, alpha-subunit complex and Arp2/3 protein complex)(Boyanova, Nilla et al. 2012).

Kinase involvement could be further validated with topological analysis performed analogously on the kinase drug targets in the platelet phosphorylation network (Figure 16). Integrative analysis of the phosphoproteome included only kinases and their direct substrates. In this case, we tested whether kinases with a high number of substrates were significantly more often targeted by drugs. We additionally identified what type of drug (approved, experimental or both) was involved. Then, kinases were tested against kinases without drug association. Overall, drug targets were enriched among well connected kinases (Figure 16A, B; mean = 15.9; p-value = 3.015×10^{-5}), consistent with results for the whole platelet interactome (Figure 14). When the two types of drugs were tested (experimental and approved), kinases affected by experimental drugs were significantly more often connected with multiple substrates (mean=14; p-value = 0.0001135) when compared to approved drug targets (mean = 13.5; p-value = 0.08372) and kinases targeted by both drug types (mean = 22.8; p-value = 0.008618) (Figure 16C). These results underline the impact of kinases with a high number of substrates as predominant targets of experimental drugs in the platelet interactome.

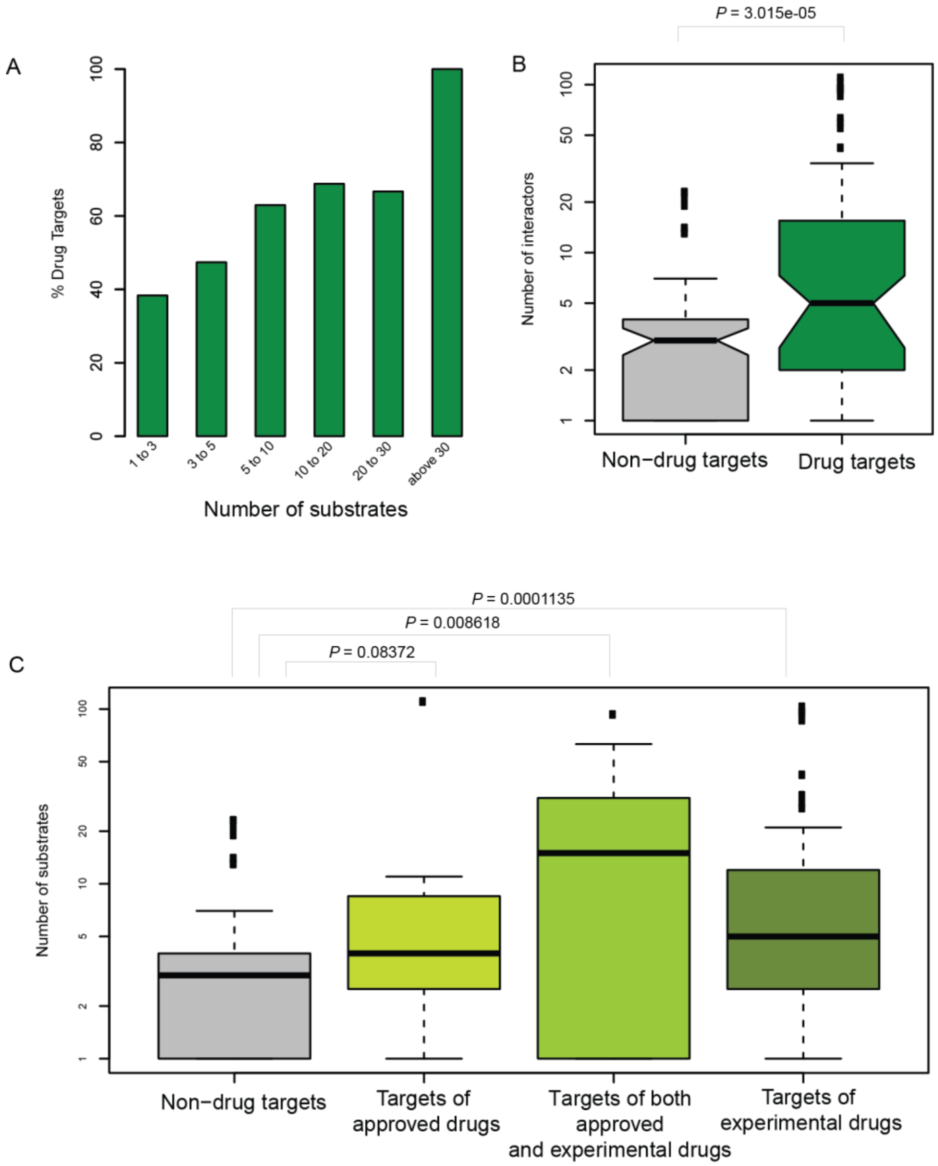


Figure 16. Kinases as drug targets

(A) Kinases were tested for enrichment of drug targets in dependence of the number of substrates. There are more drug targets found among highly connected kinases. (B) When tested for enrichment using Wilcoxon rank sum test, kinases with many substrates were significantly more often associated with drugs. (C) Kinase drug targets were also separated according to the type of drug targeting them (approved, experimental, both) and the group of experimental targets showed a high significance for drug target enrichment when compared to the other two groups(Boyanova, Nilla et al. 2012).

Investigation of drug targets was only one of the clinically relevant analyses of platelet proteins. Genetic diseases can also be mapped to platelet proteins according to the known gene information. Thus, analysis of important genes in the platelet interactome can be performed based on their association with disease. Genetic disease associations were extracted from HPRD for 701 platelet proteins from a total of 1933 human genes with available information. Similarly to the drug target proteins, disease proteins were separated into groups according to the number of their interaction partners. Subsequent topological investigation suggested that the products of these genes interact more often with other proteins and represent important hubs in the platelet network (Figure 17; p-value = 1.49×10^{-4}). When the similar analysis was performed for kinases in the phosphoproteome network, no significant increase was found with a higher number of substrates (Figure 18; mean = 10.4; p-value = 0.8836).

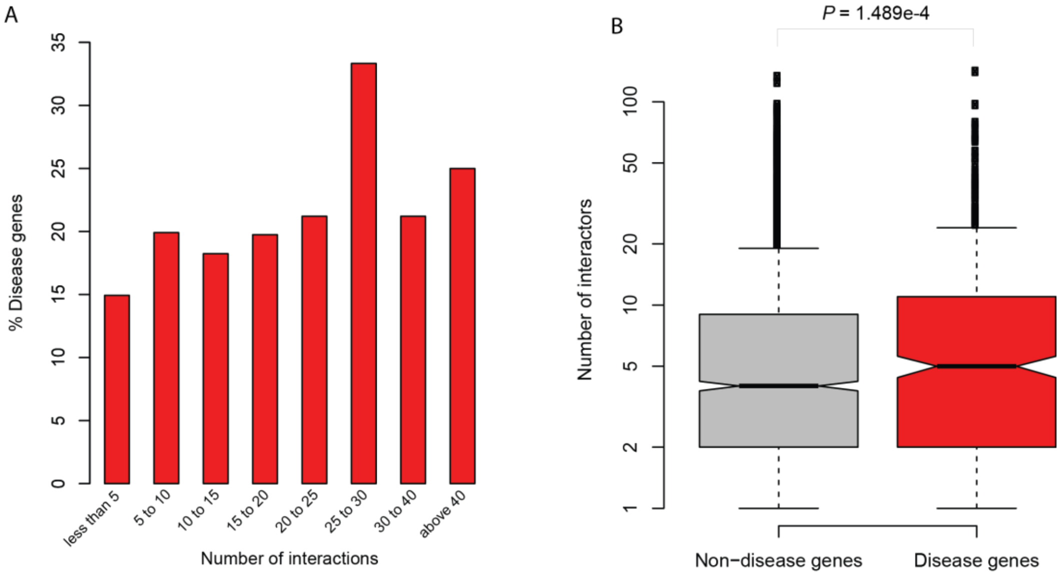


Figure 17. Dependency of platelet disease genes from their degree

(A) Proteins were grouped into intervals according to the number of their interactions. There is a higher number of disease-associated genes among well-connected proteins in the platelet interactome. (B) Datasets of disease and non-disease proteins are presented in box plots. The number of interactors is depicted on the y-axis. Their median is significantly higher in the group of disease proteins (Boyanova, Nilla et al. 2012).

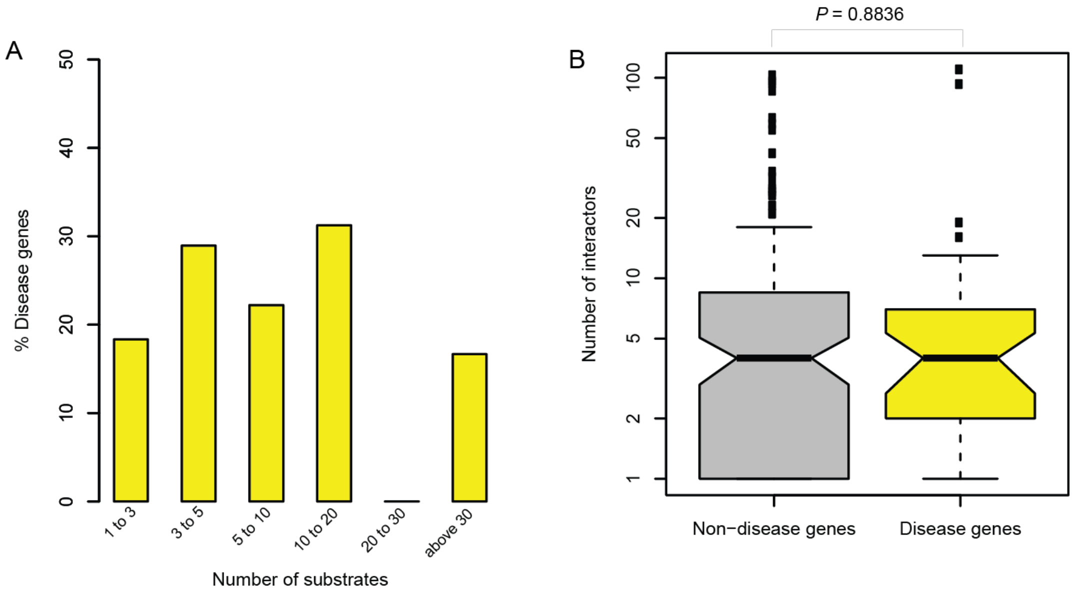


Figure 18. Dependency of platelet disease-associated kinases from the number of their substrates

(A) Kinases were grouped into intervals according to the number of their substrates. There is no change of number of disease-associated kinases among well-connected kinases in the platelet phosphoproteome. (B) Datasets of disease and non-disease kinases are presented in box plots. The number of substrates is depicted on the y-axis. Their median is not significantly changed between the two groups (Boyanova, Nilla et al. 2012).

5.2 Analysis of signal transduction in platelets

Signal transduction in platelets plays a key role in exerting platelet function. An anucleate cell can only regulate cell function through posttranslational modification of protein targets and fine-tuned signaling of downstream effectors. Therefore, systems biological investigation including network analysis can systematically pursue changes in platelets during activation and inhibition. The role of network analysis for drug development and disease biomarkers has already been discussed (Erler and Linding 2010). Understanding cellular networks has become a prerequisite to understanding complex diseases.

Platelet activation is a sophisticated process involving many specific receptors on the platelet surface. Platelet circulate in the blood flow in an inactivated state, maintained by inhibitory signal transmitted through nitric oxide (NO) secretion from the surrounding endothelium and prostaglandins, which bind to the prostaglandin receptors and cause cyclic adenosine monophosphate (cAMP) increase and platelet inhibition (Broos, Feys et al. 2011) . When a microinjury of the vessel occurs, platelets adhere to the injured surface in response to factors secreted from the exposed extracellular matrix. One of these factors, VWF, binds to GPIBA on the platelet surface and triggers transient platelet adhesion to the vessel wall (Varga-Szabo, Pleines et al. 2008). Further activation includes collagen binding to GP6 receptors, which triggers activation of PLCγ2 through complex signaling via downstream tyrosine kinases and production of diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3), which in turn lead to increase in intracellular calcium concentrations and platelet activation. Calcium concentrations in platelets are maintained either in calcium intracellular stores such as the endoplasmatic reticulum, governed by receptors controlling the Ca^{2+} influx (ITP3R) or by influx of extracellular Calcium through P2RX1 receptor (Stegner and Nieswandt 2011). Furthermore, degranulation of platelet granules and secretion of mediators such as ADP and serotonin induce further assembly of platelets to the injured site. ADP stimulation is achieved through the ADP receptors P2Y1 and P2Y12 on the platelet surface and causes additional platelet activation and aggregation. Firm platelet adhesion and aggregation is achieved by integrins, mainly by the α2bβ3 integrin, which is activated through inside-out signaling and facilitates thrombus formation by binding to fibrinogen and to other platelets through fibrinogen links (Anthis and Campbell 2011). The extraction of signaling pathways from the platelet interactome is visualized in Figure 19.

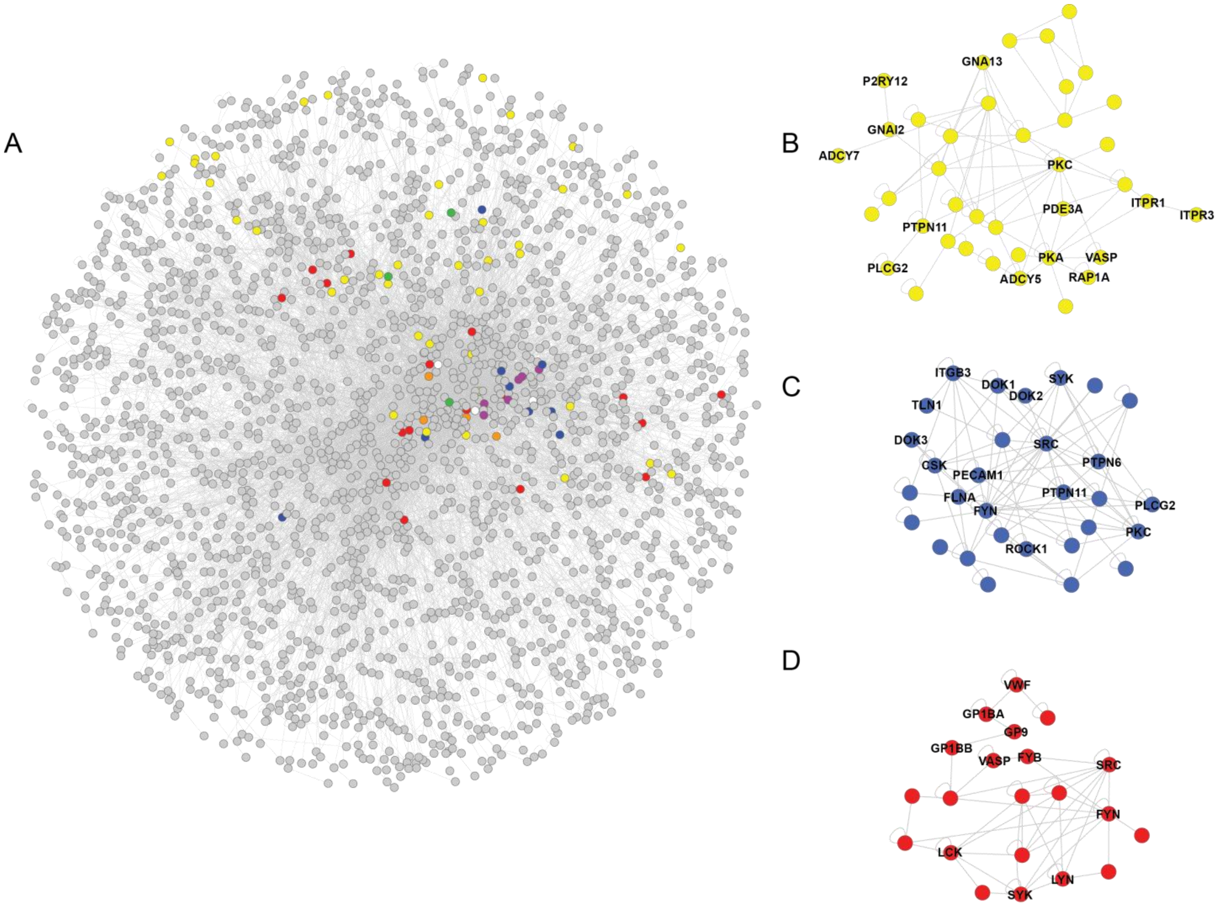


Figure 19. Signaling cascades in human platelets.

(A) Graphical representation of three main platelet signaling pathways: vWF signaling (red), integrin signaling (blue) and ADP signaling (yellow), platelet proteins (grey). The common proteins for two or more pathways are shown with different colors - vWF_integrin (violet), vWF_ADP (orange), ADP_integrin (green), ADP_vWF_integrin (white). For illustration of subnetworks we show the topology and key nodes of the following subnetworks: (B) subnetwork, of ADP signaling (C) subnetwork of the integrin signaling pathway and (D) subnetwork containing vWF-signaling proteins (Boyanova, Nilla et al. 2012).

Combining knowledge from literature phosphorylations, interactions and drug data and disease associations allows integrated network analysis in platelets based on experimentally validated data. By using this approach, I created a network model of ADP signaling to complement a developed Boolean ADP model by Mischnik et al. Furthermore, I analyzed integrin signaling in platelets in the context of experimentally validated phosphorylations from a collaborating group (Zahedi, Lewandrowski et al. 2008) and developed a new hypothesis of integrin inside-out signaling including a newly identified phosphorylation site (Ser^{269}) on the integrin inhibiting docking protein 1 (DOK1). In a further approach, I used integrated network analysis to reveal the network context of experimentally measured SH2 domain binding proteins and elucidate their role during platelet activation.

5.2.1 Modeling ADP signaling

In this part of the analysis, I focused on platelet activation and inhibition during thrombosis. Drug development is mainly concentrated on the tight balance between platelet activatory and inhibitory signals and up-to-date there have been successful applications of combinational drugs acting on these pathways. Besides the cyclooxygenase 1 and 2 (COX1/2) inhibition and attenuation of TXA2 formation by acetylsalicylic acid (aspirin) (Patrono and Rocca 2012), receptor inhibitors such as Clopidogrel (inhibitor of the P2RY12 receptor) (Dorsam, Murugappan et al. 2003) and Ridogrel (inhibitor of the thromboxane receptor and thromboxane synthesis inhibitor) (Heinisch, Holzer et al. 1996) have been administered successfully for reduction of atherosclerotic events and thrombolysis. Nonetheless, antithrombotic therapies still need improvement mainly for reducing the risks of drug side effects. Therefore, modeling some of the most important pathways can considerably increase the chance of finding novel targets for drug development.

One of the main activatory pathways in platelets, which have been the focus of pharmacological development, is the adenosine diphosphate (ADP) signaling pathway. ADP acts as an autocrin ligand for platelets, which is released upon platelet binding to injured endothelium of blood vessels, where the extracellular matrix has been exposed. ADP as signaling trigger induces only weak reversible effects. Nonetheless, it is a crucial second messenger in platelet signaling when released from dense granules in the platelet at high concentrations (Konig, Nimtz et al. 2012).

P2 receptors in platelets are the main targets for ADP and ATP. While P2Y1 and P2Y12 bind to ADP, P2X1 is a receptor for ATP (Cattaneo and Gachet 2001). I will focus here mainly on the ADP receptors and their downstream signal transduction. The P2Y1 receptor is coupled to a Gq protein, which activates PLCβ in platelets, and thereby mediates the mobilization of ionized calcium from intracellular stores, eventually facilitating platelet shape change and reversible aggregation (Figure 20). The P2Y212 receptor is coupled to adenylate cyclase through an inhibitory Gi protein and triggers ADP-induced aggregation without former shape change. Additionally, this receptor mediates downstream events, which eventually lead to the upregulation of PI3K and platelet secretion from α- and dense granules. Full platelet aggregation response after ADP stimulation can only be achieved with a combined action of P2Y1 and P2RY12 receptors (Liu, Pestina et al. 2004).

Reduction of cAMP levels facilitates inhibition of PKA, as cAMP binds to the regulatory subunits of type I and II PKA, thus activating the kinase. The activity of PKA in platelets causes downstream phosphorylation of protein targets and inhibition of platelets in general. These inhibitory effects are regulated by cAMP-phosphodiesterases, which degrade cAMP to 5'-AMP, thereby attenuating effects of agonist increasing cAMP levels (Feijge, Ansink et al. 2004).

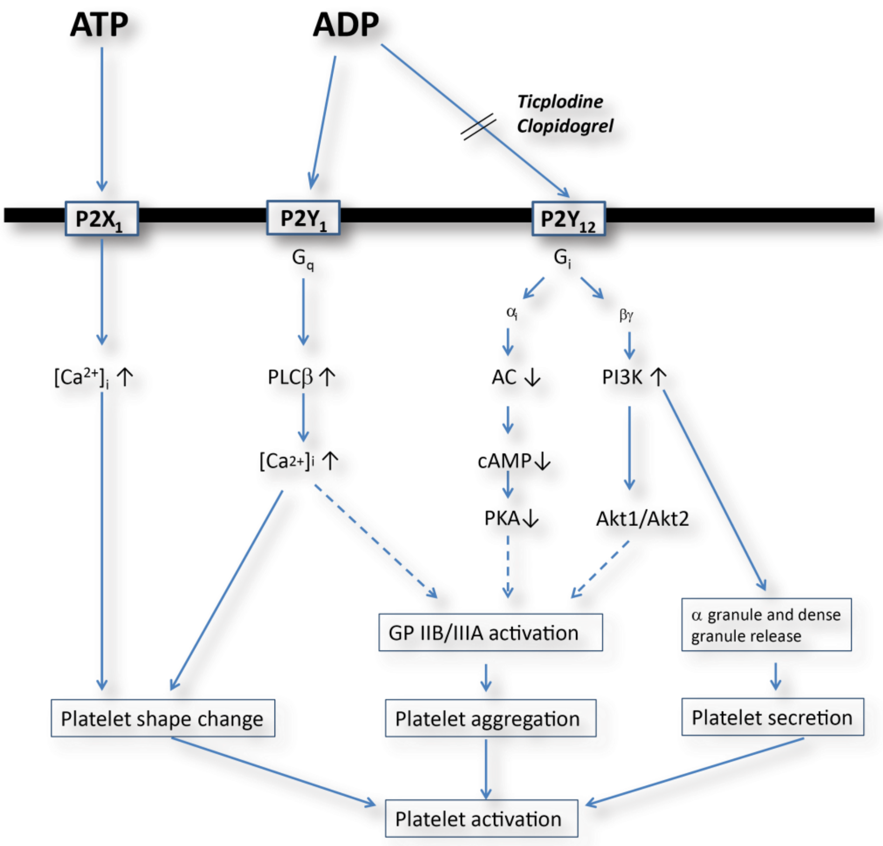


Figure 20. Signaling of P2 receptors in platelets

The figure represents a scheme of downstream receptor signaling of P2 receptors in platelets. The P2X1 receptor, which binds to ATP, causes a quick influx of extracellular Calcium, leading to platelet shape change without platelet activation. P2Y1 is coupled to a Gq protein, which mediates the activation of the membrane bound PLCβ, followed by an increase in Calcium concentration due to Calcium release from the intracellular stores. The calcium increase causes platelet shape change and activation of GPIIB/IIIA inside-out activation after a number of reactions. P2Y12 receptor is responsible for inducing platelet aggregation, it is coupled to a Gi protein, which facilitates platelet aggregation by reducing the cAMP level in platelets and thus preventing the inhibitory action of PKA. On the other hand, PI3K is activated, which ultimately leads to platelet secretion after granule release and activation of firm platelet adhesion and aggregation by inducing Akt kinase activity. Inhibitors of P2Y12 receptor (Ticlopidine and Clopidogrel) are also shown. Figure adapted with modifications from (Joo 2012).

Here, we present a model for ADP signaling in platelets, enriched with information from literature on human and platelet phosphorylation events, interactions as well as kinase predictions. The network analysis and network model was part of a bigger project for generating a Boolean Model of platelet ADP-dependent activation. This model aims to demonstrate how receptor signals are carried out to process activation information. Furthermore, the model shows different steps of platelet activation and potential threshold behavior in accordance with experimental observations. Using Boolean logic, the ADP signaling model is presented in four different phases of activation. Each of these phases can be visualized as a network, including phosphorylation and interaction information extracted from literature studies. A detailed analysis of these networks then allows better understanding of observed changes during activation. The main network model is shown in Figure 21.

According to the model, P2 receptors (P2RY1 and P2RY12) transduce ADP signal to coupled G-proteins (e.g. the inhibitory Gi protein GNAI2 for P2RY12), which then further inhibit adenylate cyclase function and thus reduce cAMP levels in platelets. Central kinases involved in downstream signaling such as the activatory kinases Src and PKC and the inhibitory kinase PKA can be further analyzed in a network context along with their phosphorylated substrates. The downstream effects of platelet activation such as shape change, triggered by the activation of integrins, are also included in the model. Phosphorylations were extracted from literature and additional linking proteins were added to maintain the network integrity. Thus, the network represents phosphorylation events, which were indeed measured in the human or platelet system and creates a realistic representation of the signal flow during ADP signaling.

In a second approach, the activation level of each protein during the four Boolean model phases was mapped to the ADP signaling network and visualized with different colors. Furthermore, interesting kinase-substrate relationships of the proteins with changed activation level were extracted from the network and analyzed in detail according to the experimental information available in literature (data not shown).

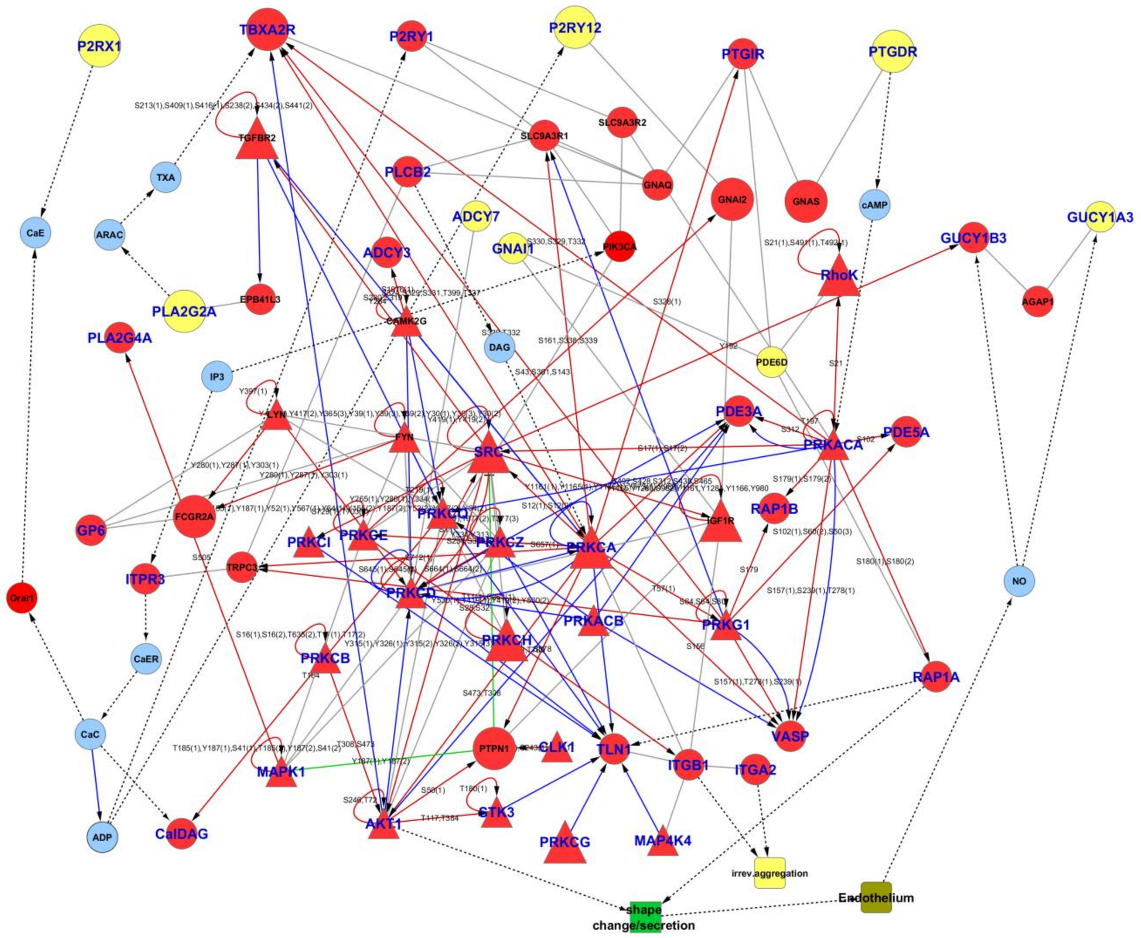


Figure 21. ADP signaling model

The network represents ADP signaling starting from the receptors and following the signal flow to the changes induced in platelets such as shape change. All shown interactions (grey lines), site-specific phosphorylation events (red arrows with labeled sites) and dephosphorylation events (green lines) were extracted from literature studies based on the *PlateletWeb* knowledge base. Kinase predictions (blue lines) were additionally added to the model. Kinases are presented as triangles and proteins as circles. Phosphorylated proteins are depicted in red and proteins from the initial Boolean model network are shown with blue labels. The rest of the proteins were added for maintaining the network structure and improving the visualization of the signal flow (for example G-proteins). Metabolites and small molecules such as NO and Ca^{2+} are presented in light blue circles. A large size of the nodes denotes that the protein is associated with a genetic disease. Following the signal from receptors to phosphorylation events downstream, triggered by the activity of key kinases such as SRC and PKC, the model presents the release of calcium flux from the endoplasmatic reticulum and associated changes of a number of protein targets such as RAB1, integrins (e.g.ITGA2), CalDAG and Talin, which are crucial for irreversible aggregation caused by inside-out activation of the ITGB3 integrin.

5.2.2 SH2 domain binding proteins

The tight balance of platelet activation can only be achieved by a very precise signaling regulation. Signaling during platelet activation is mainly triggered through the interaction of receptors with their ligands and the downstream transduction of the signal flow through second messengers and signaling protein domains.

The SH2 domains, Src homology 2 domains, are important signal transducers as they bind to tyrosine phosphorylated proteins and thus mediate signal responses. Protein tyrosine kinases (PTKs) and their substrates play a critical role in the regulation of cell processes such as proliferation, differentiation, movement and immune responses, as well as pathological conditions such as cancer (Hunter 2000). Tyrosine kinases also play a major role in platelet activation by phosphorylating a large number of substrates (222 platelet substrates (Boyanova, Nilla et al. 2012)), ultimately leading to platelet shape change, aggregation and thrombus formation. SH2 domains were found enriched among platelet proteins (p-value = 4,73 x 10^{-5})(Boyanova, Nilla et al. 2012), which furthermore indicates their importance in signaling processes. Therefore, a closer look on SH2 domains can elucidate how platelets initiate the activating response.

In a previous study, Machida et al developed a method which identifies tyrosine phosphorylation sites by the use of SH2 domains and a far-western blot technique (Machida, Thompson et al. 2007). Thus, the global tyrosine phosphorylation state of the cell can be analyzed with a single experiment. They further extended their approach to platelet proteins and obtained tyrosine phosphorylations for 19 proteins binding to two separate SH2 domains (EAT2 and ABL2).

Working in collaboration with this group and their results on the platelet response, I focused on the network analysis of the identified SH2 domain targets in the context of ADP signaling. At first, the size of the obtained protein bands was analyzed to assign them to known proteins. Then, a bioinformatical step was added including the determination and validation of the identified proteins according to their molecular weight, an interactome analysis of the detected SH2 domain binding proteins and their position in the ADP signaling network. The experimental design was predefined to identify proteins with tyrosine phosphorylation, but the responsible kinase couldn't be determined using far-western techniques. By the integration of kinase data from the

PlateletWeb database, coupled with site-specific phosphorylations and dephosphorylations, interactions and platelete-specific information, the network around the SH2-binding proteins could be constructed and visualized along with the ADP signaling pathway, explained in detail in the previous chapter (5.2.1 ADP signaling). As the exact tyrosine phosphorylation sites were not yet determined in the lab experiment, phosphosites from known literature sources were added to the signaling network, thus giving indication about the phosphosite responsible for binding of the SH2 domains to these proteins. Figure 22 summarizes all available kinase and site-specific information in a single network.

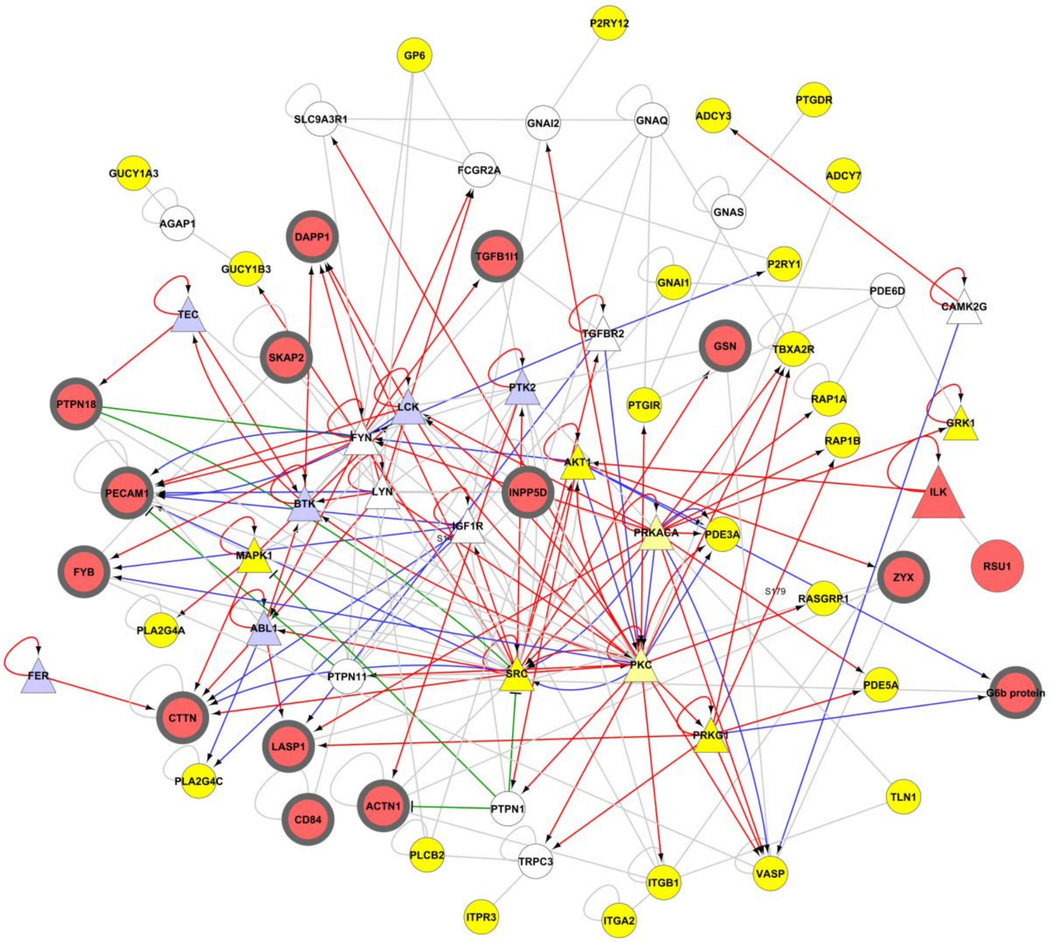


Figure 22. ADP model with SH2 domain proteins

The figure represents the 16 SH2 binding proteins in the context of ADP signaling. **Yellow**: ADP model proteins, **Red large circle**: SH2 domain binding proteins, **Blue**: added kinases which target SH2 binding proteins, **White**:

additional proteins which hold the ADP model network structure, **Grey line**: protein interaction, **Red arrow**: literature-derived phosphorylation, **Blue arrow**: kinase prediction for experimentally-validated phosphorylation site in platelets, **Green line**: literature-derived dephosphorylation ,**Circle**: protein, **Triangle**: kinase, **Thick circle line**: the protein is phosphorylated on a tyrosine residue according to literature studies

Results reveal well-known signaling players during platelet activation, such as PECAM-1 and FYB. Notably, 14 out of the 16 SH2-binding proteins have a known tyrosine phosphorylation in literature studies, which provides additional confirmation for the indicated tyrosine phosphorylation of these proteins. Kinase-substrate data extracted from the *PlateletWeb* knowledgebase revealed six tyrosine kinases, known to phosphorylate some of the identified proteins (LCK, BTK, FER, ABL1, TEC, PTK2). The Bruton tyrosine kinase (BTK) mediates platelet responses during the initial steps of platelet adhesion and activation, which takes place after binding of platelets to vWF on the injured surface (Liu, Fitzgerald et al. 2006). Analogously to BTK, the Src family kinases FYN and LYN phosphorylates downstream targets and thus mediates platelet responses triggering platelet activation and integrin activation (Yin, Liu et al. 2008). The resulting network, integrating multiple sources of information, not only visualizes the functional network context of the SH2-binding proteins identified in the study, but also allows kinase predictions to be added to the network. The intricate balance of the platelet activation state can therefore be analyzed in a global interaction network context, which sheds light on possible phosphorylation events, already described in literature. The site-specific annotations can be further extracted and used for analysis of platelet tyrosine substrates. For example, the phosphorylation of PECAM-1 by LCK is well described in literature (Newman, Hoffman et al. 2002).

This analysis serves as a first step towards a full integrative approach for the investigation of SH2 binding protein studies in the future. Nonetheless, the interpretation is only limited due to the lack of substantial evidence for these candidate proteins. To achieve a full and comprehensive network analysis, the identified candidates must be further analyzed using mass spectrometry to ensure that the mapping of the western blot bands can be assigned to these proteins in particular.

5.2.3 DOK1 phosphorylation and integrin activation

Integrin activation is an important later step in platelet activation, because it facilitates firm adhesion and thrombus formation (Varga-Szabo, Pleines et al. 2008). Integrins are bidirectional molecules which require a conformational change for achieving their active state. This process, called inside-out signaling, is mediated by various signaling cascades inside the activated platelet. Thrombus formation is mainly carried out by the integrin α2bβ3 which is activated from the inside by a number of signaling cascades ultimately activating the protein talin (Figure 23). Activation of platelets causes an increase of intracellular Ca2+ and DAG, both stimulating the activity of diacylglycerol regulated guanine nucleotide exchange factor I (CalDAG-GEFI) and PKC. Both signaling paths cause the activation of Rap1b to Rap1-ATP and its translocation to the platelet membrane using the Rap1-GTP-interacting adaptor molecule (RIAM). Then the signal is transmitted to talin, an integrin activating molecule. When talin binds to the β3 tail of the integrin via its FERM domain, a conformational change takes place and the integrin gains the ability to bind fibrinogen and thus connect platelet cells over fibrinogen bridges (Wegener, Partridge et al. 2007). There are various mechanisms for integrin regulation in platelets, and it is known that a docking molecule DOK1, a protein containing a pleckstrin domain (PH) and a phosphotyrosine binding (PTB) domain (Songyang, Yamanashi et al. 2001), competitively binds to the integrin β3 tail and thus hinders binding of talin and activation of the receptor. There have been both in vivo and in vitro evidence for DOK1 binding to the cytoplasmatic tail of β3 integrin (ITGB3)(Calderwood, Fujioka et al. 2003). The affinity of β3 increases when the tail is phosphorylated on the site Y773, where both DOK1 and talin bind competitively with their domains (Figure 23, Figure 25).

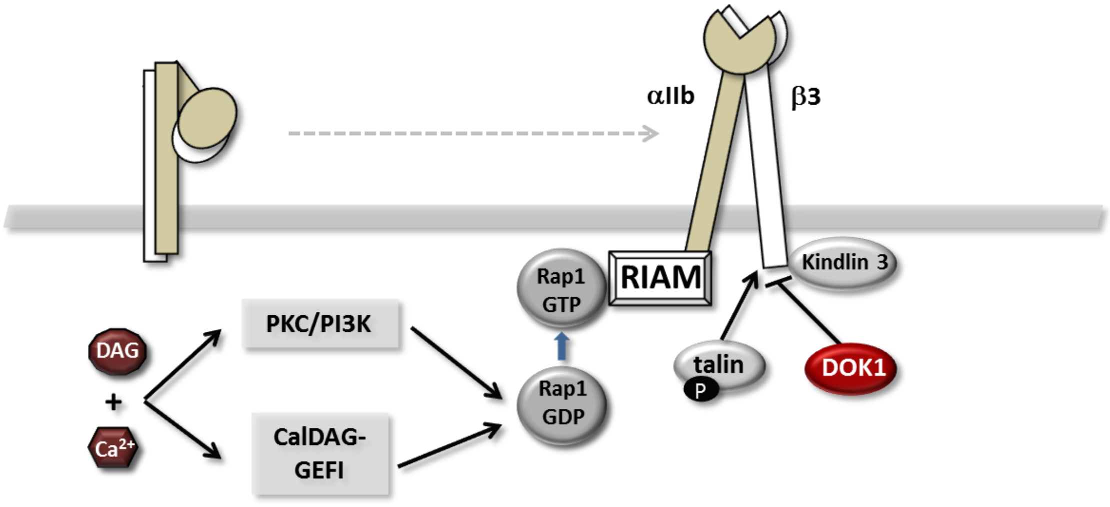


Figure 23. Integrin inside-out signaling

The activation of integrin α2bβ3is achieved by signaling triggered from inside-out. When levels of Ca2+ and DAG are increased, this causes the stimulation of CalDAG-GEFI and PKC/PI3K activity which in turn activate Rap1.
An "activation complex" is formed, which contains the proteins Rap1, RIAM and talin and causes cytoskeletal rearrangements along with conformational changes in the integrin molecules. The bent inactive form of the integrin turns into the activated form with exposed fibrinogen binding site. The exact mechanisms of kindlin 3 has not yet been described. The figure is modified from (Broos, Feys et al. 2011).

Therefore, focusing on this particular regulation and following the indication of the pathway enrichment in platelets, where integrin signaling was overrepresented, we used the *PlateletWeb* knowledge base options for integrated analysis of the proteins from this pathway using combined information from the platelet interactome, phosphorylation signaling network functional and drug data. From the entire platelet network we extracted a subnetwork of integrin signaling near the α2bβ3 receptor (Figure 24). We generated an overview of signaling events in the integrin pathway including literature phosphorylations and kinase predictions (Linding, Jensen et al. 2007; Miller, Jensen et al. 2008) for experimental phosphorylation sites measured in platelets (Zahedi, Lewandrowski et al. 2008) (Figure 24; supplemental Tables 4 and 5). The predicted kinase-substrate relationships in the network were supported by experiments in other human cells such as PKA phosphorylating SRC at Ser^{17} (Obara, Labudda et al. 2004) and SRC phosphorylating ITGB3 at Tyr^{773} (Datta, Huber et al. 2002) (same as Tyr^{747}, new nomenclature from HPRD) (Figure 24). Furthermore, the assembled network contains information on associated drugs, most of which are in the stage of development (noted as experimental drugs).

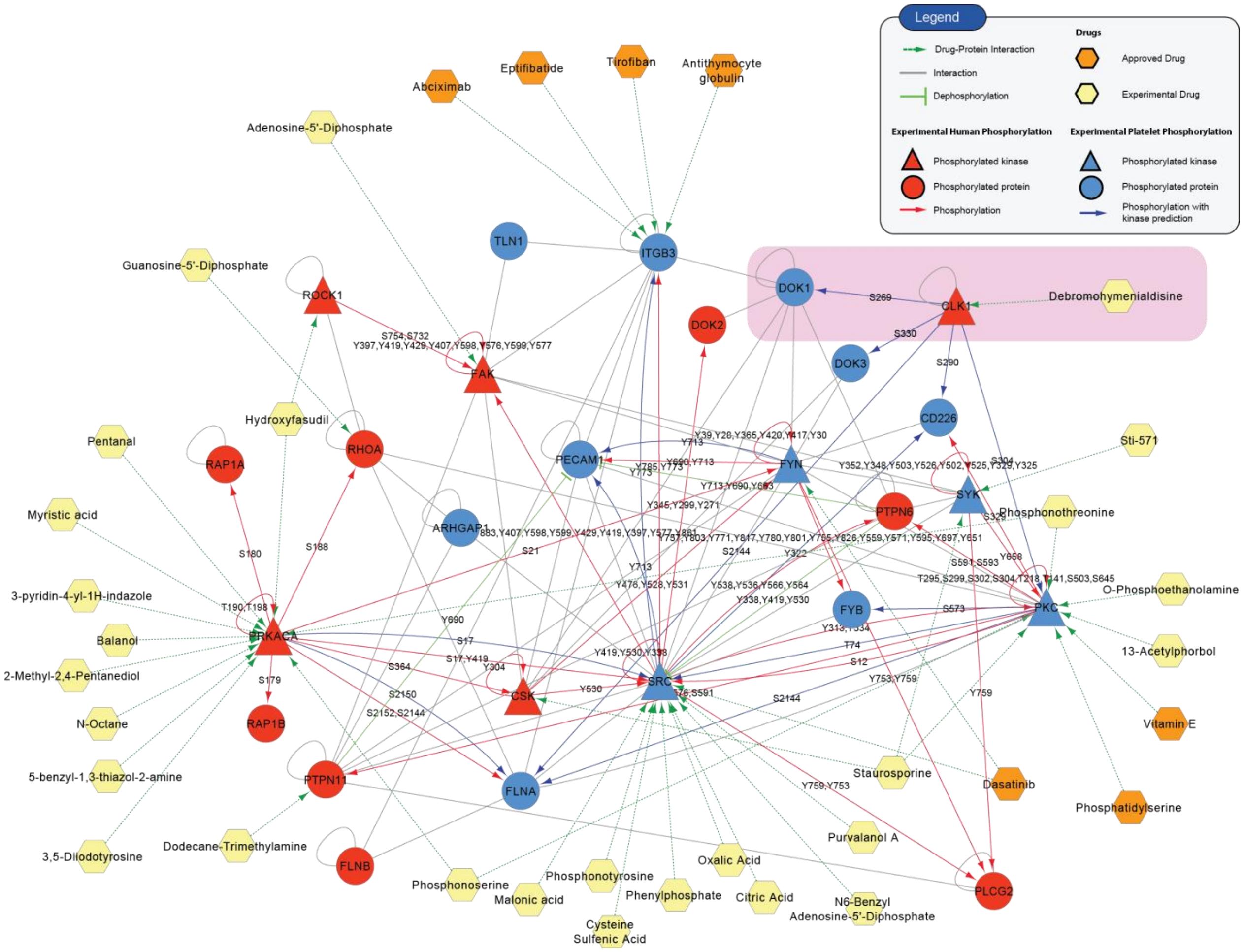


Figure 24. Integrin signaling pathway with focus on DOK1 phosphorylation

Visualization of the core integrin signaling created by integrating information on phoshorylations, dephosphorylations (green lines) and interactions (grey lines). *P*hosphorylations are depicted according to their source of detection: red arrows indicate phosphorylations reported from human cells (HPRD), blue arrows are used in the cases where a kinase prediction is assigned to an experimentally-validated phosphorylation site. The protein nodes are coloured according to the source of phosphorylation (red: phosphorylated in human cells, blue: phosphorylated in platelets, yellow: a platelet non-phosphorylated protein*)* and the phosphorylation site is presented on each directed edge. Drugs are visualized with different colours according to the type of drug: investigational (experimental) or approved. DOK1, a docking protein associated with integrin β3 (ITGB3) binding, is phosphorylated on Ser[269] (highlighted region). By further integrating kinase prediction and drug target information, the platelet kinase CLK1 has been proposed and a putative therapeutical approach using the inhibitor debromohymenialdisine can be suggested. This figure was originally published in Blood (Boyanova, Nilla et al. 2012).

During the network analysis, we focused on the direct interactors of the β3 integrin tail (ITGB3), a yet functionally uncharacterized phosphorylation site was detected in unstimulated human platelets at Ser^{269} of DOK1. The Ser^{269} phosphorylation site is in close proximity to the IRS-type PTB domain, which facilitates binding of DOK1 to ITGB3 (NPLY motif, Figure 25) and inhibits its activation (Oxley, Anthis et al. 2008). The integrin activating molecule talin competitively binds to the same motif, which suggests that the phosphorylation site might influence the balance between DOK1 and talin binding and thus regulate integrin activation.

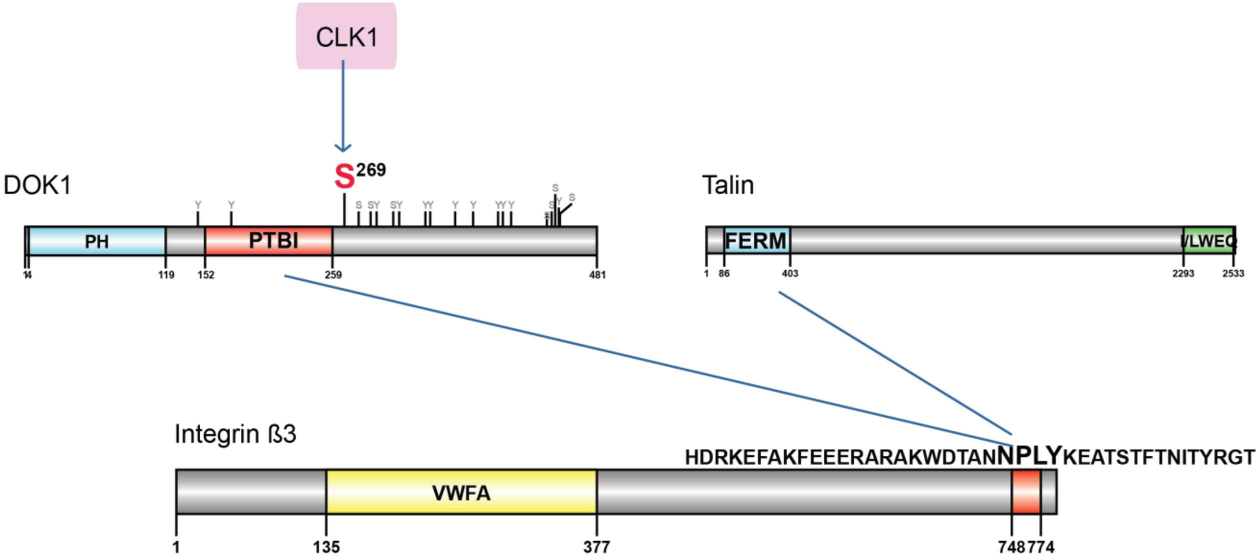


Figure 25. Structure of DOK1, Talin and Integrin β3 and their interaction

Schematic representation of the DOK1 Ser^{269} phosphorylation site and the competitive binding between DOK1 and talin for the NPLY-motif of the integrin β3-tail. DOK1 binds to the integrin and prevents talin from binding and activating the receptor. *This figure was originally published in Blood (Boyanova, Nilla et al. 2012).*

By integrating both kinase prediction and drug data, hypothesis for analysis of the functional role of DOK1 phosphorylation could be generated. The applied bioinformatical prediction method (Linding, Jensen et al. 2007; Miller, Jensen et al. 2008) identified CLK1 as the kinase responsible for phosphorylating Ser^{269} on DOK1. CLK1 is a kinase found in megakaryocytes, proplatelets and platelets (Schwertz, Tolley et al. 2006). CLK1-dependent splicing of tissue factor (TF) pre-mRNA in platelets is a previously unrecognized pathway to fibrin formation and stabilisation of a platelet thrombus (Schwertz, Tolley et al. 2006). According to the associated drug data, CLK1 is inhibited by the chemical compound Debromohymenialdisine (Ulitsky and Shamir 2007) (a marine sponge alkaloid), which has been isolated from Axinella sp. (Song, Qu et al. 2011). This compound was reported to act as cyclin-dependant kinase (CDK) modulator with pharmacological activities for treating osteoarthritis and Alzheimer’s disease (Roy and Sausville 2001). Thus, we propose a possible mechanism of investigating the functional role of DOK1.

5.3 Identification of protein functional modules using functional interaction scores

Many challenges have emerged in recent years for data analysis of large-scale proteomic datasets. The mere generation of gene lists is not enough to explain systematic effects and changes in signaling regulation. Therefore, network analysis uses the extraction of functional modules from a set of proteins, revealing information about the main functions regulated in the analyzed sample. The term “functional module” has first been presented in gene expression analysis, where functional modules of differentially expressed genes can be investigated for network regulation. The problem of transferring this idea to proteomics is that proteomics provides mostly qualitative and insufficient results for a large systems biological analysis. The large number of identified proteins makes it hard to isolate functional modules, therefore new approaches are needed. Including information for protein-protein interactions can be quite useful but insufficient to optimally solve the problem of identifying functional modules in the network and their connections with each other and the rest of the network. Additional functional information on the interacting proteins can help to focus on the connecting paths with more experimental confirmation and better biological interpretation of the network. Functional association information on the edges of the human interactome using interaction values can be calculated according to the functional similarity of the interacting proteins. Gene Ontology (GO) is a hierarchical structure containing functional terms, grouped in ontologies (Ashburner, Ball et al. 2000). There are three main ontologies (branches) of the hierarchical tree: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC) containing a set of terms with increasing specificity towards the lower branches of the tree. Proteins are assigned to specific biological terms according to their functionality in an either manual or automatic way. The comparison between two genes is therefore possible using data from their functional annotations. The first step for quantifying the similarity between two genes is the calculation of semantic similarity scores based on the similarity of their annotated terms. These scores can be provided by pairwise or group-wise measures according to the type of comparison used. There have been various approaches in recent years, aiming at optimizing the semantic similarity of GO for an improved biological interpretation (Lord, Stevens et al. 2003; Jain and Bader 2010; Ramirez, Lawyer et al. 2012), but only few of these measures are considered best by Guzzi et al (Guzzi, Mina et al. 2011): Resnik (Resnik 1995), simGIC (Pesquita, Faria et al. 2008), simIC (Li, Luo et al. 2010) and TCSS (Jain and Bader 2010). Pairwise strategies comparing pairs of GO terms can't directly be applied to genes and proteins, therefore mixing stragies for the transformation of all pairwise term similarities into a single value are needed. Such strategies have been introduced, using different options of the GO graph: average (Lord, Stevens et al. 2003), maximum (Sevilla, Segura et al. 2005), Best Match Average (Azuaje, Wang et al. 2005), funSim (Schlicker, Domingues et al. 2006), Information Theory-based Semantic Similarity (Tao, Sam et al. 2007), FuSSiMeG (del Pozo, Pazos et al. 2008). These approaches were further modified for ranking candidate disease genes based on the comparison of their functional annotations (Schlicker, Lengauer et al. 2010). When the semantic similarity information is transferred to a PPI network, proteins with high functional similarity are expected to be found in similar cellular processes and an interaction between them is therefore most probable. The edge score of the interactome would then reflect functional relevance of the particular interaction in the network.

During the following analysis, we developed a new score based on the score by Schlicker et al (Schlicker, Domingues et al. 2006), assigning functional information to the edges of the human interactome. Using a previously developed functional module detection algoritihm (Dittrich, Klau et al. 2008) now enhanced with functional information as edge values, various cell lines were characterized functionally to test how well the algorithm performs. I was mainly involved in validating the accuracy of the algorithm along various small and large proteomic datasets and biological interpretation of the obtained results.

5.3.1 Concept of functional interaction scores

Network analysis has become important in the field of transcriptomics and proteomics due to the search for functional information and relationship data in complex proteome datasets, which still remains elusive in many cases (Pieroni, de la Fuente van Bentem et al. 2008). Integration of multiple data sources into a single network context, also described as integrated network analysis, combines expression data with PPI information and helps to identify modules of genes with similar regulation (functional modules). Developed at first in the field or transcriptomics, this type of analysis is also gaining importance in proteomics due to its powerful methods of systems biological interpretation of the obtained data. Proteomics still suffers from many limitations such as poor data coverage and consistency, reproducibility and limited detection of low-abundance proteins or proteins from specific subcompartments, such as membrane proteins. Especially in the case of qualitative proteomics, where only the presence or absence of proteins is examined, there is a great need for methodologies, which allow the network analysis of obtained datasets. Integrated network analysis can be useful to overcome many of the challenges faced by proteomics today (Goh, Lee et al. 2012). Biological networks have modules with different functionality (Hartwell, Hopfield et al. 1999). The components of the human interactome are highly connected and there are often more than one ways to connect the sample proteins. A method for investigating the optimal protein module from a biological sample can greatly improve network analysis in proteomics. One such approach is an algorithm for exact functional decomposition of large networks in functional modules based on scoring of nodes and finding the optimal maximum-scoring subgraph (Dittrich, Klau et al. 2008). It was initially developed to analyze microarray data and to extract optimal subnetworks using scores based on signal-noise decomposition from a node-weighted network. Originally, the expression values of the genes from a microarray experiment are transformed into p-values and these p-values are separated into a signal and noise component using a beta-uniform mixture model (BUM). Then, the problem of finding maximum weighted subnetworks is transformed into the well-known prize-collecting Steiner tree problem (PCST). Using this module detection algorithm, functional modules of similarly regulated genes could be extracted representing different tumor subtypes.

In this thesis, the algorithm was transferred to proteomics data and complemented with functional similarity between interacting proteins as edge scores in the network. Nodes from the protein sample were given a positive score, while the rest of the interactome was assigned a negative score. The optimal solution extracted from the algorithm then consists of the maximum positive scoring nodes (proteins) and interactions connecting them. The modules are not unique because there are many possibilities to connect the positive nodes into a maximum-scoring subgraph. Therefore, a weight on the edges would help to focus the resulting network onto paths with higher probability, mainly due to the functional resemblance of connected proteins in the solution.

We propose a new method for investigating large proteomic datasets by weighting nodes and edges of the investigated network according to the functional information of the interacting proteins in the PPI network extracted from Gene Ontology (Figure 26). The Gene Ontology (GO) is a hierarchical structure, which annotates genes according to their functionality in three main categories: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC) (Ashburner, Ball et al. 2000). The functional annotations are represented in a directed acyclic graph (DAG) with specificity of terms increasing towards the bottom of each branch. A prerequisite to introducing functional interaction scores is the availability of a functional information score, which measures the similarity of two proteins in terms of their functional annotation in GO - the GO semantic similarity score (Guzzi, Mina et al. 2011). The rationale behind a scoring system based on GO is that proteins with a high functional similarity will be involved in similar cellular processes, thus the interaction between them will be given a higher priority. In this case, the edge score would reflect functional relevance of the particular edge in the PPI network.

In our method, proteins identified in a sample were mapped to a previously defined interactome network (10688 nodes and 55196 edges) and the existing interactions between them were weighted with a score based on a previously developed GO semantic similarity score by Schlicker et al (Schlicker, Domingues et al. 2006), which we adjusted by adapting it to a network context. We will further refer to these scores as functional interaction scores as they represent the functional similarity between each interacting pair. We assigned proteins from the biological sample positive values derived from the functional interactions scores (Experimental Procedures), whereas linking proteins from the interactome (not present in the proteome sample) were given the average of all interaction scores (negative score) (Figure 26B). Based on the interaction and protein scores, the algorithm searches for the maximally scoring subgraph, thus identifying the most likely module of functionally similar proteins (Figure 26A). Highly scoring interactions connect proteins with a high functional resemblance, thereby improving their chance of being included in the extracted module. Using the functional interaction scores of interacting protein pairs ensures that the algorithm includes paths having high similarity over those having low similarity (Figure 26C). Linking proteins from the human interactome are included only if this increases the maximum score of the resulting subnetwork. These additional proteins might be important for signal transduction in the analyzed network and can be potentially present in vivo. As proteomic studies are often fractionated and detection of some proteins proves to be extremely difficult, this approach is useful for unraveling the network context of identified proteins along with proteins missing in the original sample due to technical difficulties.

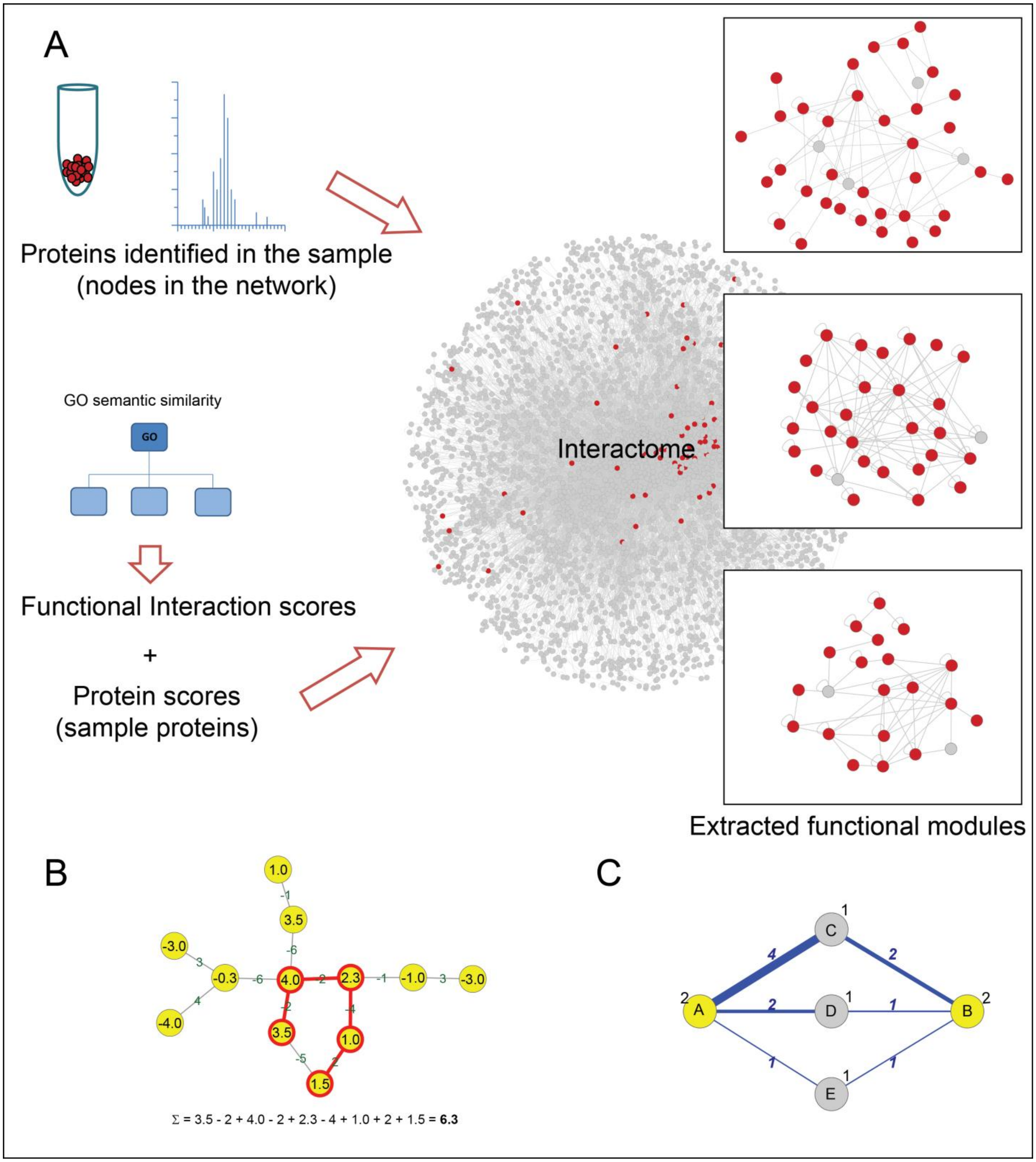


Figure 26. Concept of module detection using functional similarity

(A)A flowchart of the subnetwork extraction algorithm enhanced with functional information. Proteins identified in a proteomic sample by mass spectrometry are mapped to the human interactome. The human PPI network is given functional interaction scores based on GO semantic similarity of the interacting genes and protein scores (node scores) based on the interaction scores. Clusters of functionally similar genes can be extracted using the module detection algorithm (B) The calculation of protein scores is presented in a simplified example. Proteins identified in the sample are given a score based on the number of adjacent edges, while proteins from the rest of the interactome are assigned the negative average value of all functional scores in the network. In the given example the score of each node is obtained as the negative average of the scores of all adjacent edges and the proteins forming the maximum-scoring subnetwork are highlighted with red borders. (C) An example for the benefits of inducing GO semantic similarity into the network analysis. From equally scoring graphs, preference is given to the one with a higher overall score, thus choosing the one including higher semantic similarity paths. Functionally similar paths are so chosen over other possible solutions. Proteins A and B are identified in the sample and connected over the linking protein C, because the functional similarity between each of the sample proteins with C is higher than any other connecting path.

5.3.2 Extraction of network modules using qualitative proteome data

Proteins obtained with proteomics analysis can be visualized in a network context, even though they are not directly connected with each other. Integrating functional interaction scores to the module detection algorithm helps to connect these proteins in an optimal way based on the functional context of the protein modules and the scores of the interactions connecting them.

In order to test this, we applied the algorithm on a small biological sample from a human gastric cell line (AGC, originating from a gastric epithelial cell line) analyzed after infection with avian influenza virus (H9N2) using mass spectrometry. The study identified 22 proteins (Liu, Song et al. 2008). We performed two separate analyses of this protein sample to investigate the advantages of functional interactions scores: including and not including interaction scores in the module extracting algorithm (see Materials and Methods). When adding functional interaction scores, we used the previously described method for obtaining functional interaction scores (Materials and Methods) and assigned proteins from the sample a node score derived from the edge score (Figure 27B).

As the functional interaction scores are based on GO annotations, which are extracted from literature information, we expected the algorithm including functional interaction scores to derive functional modules with higher functional similarity and higher number of publications available for each interaction. Thus, information of the edges is based on well-studied interactions, which are traceable experimentally and can serve as a solid basis for hypothesis generation after integrated network analysis.

The method incorporating functional interaction scores (Figure 27B) yielded 39 proteins, whereas the method without interaction scores extracted 37 proteins (Figure 27A). Both networks contained linking proteins and interactions, which partially overlap (14 interactions and 5 linking proteins are the same). There were 28 proteins in common for the two solutions along with unique proteins, which appeared only when functional scores were introduced to the edges (AKT1, EPB41, GNG4, HRAS, HSPA4, JAK2, KRT5, PDK1, RAF1, YWHAB). The number of literature citations for each interaction was very low in the network without functional scores (Figure 27A). In the network with scores there were five interactions with more than one citation (Figure 27B), indicating the algorithm chooses well-annotated interactions.

It was important to examine in detail the exact paths chosen by the algorithm and whether the addition of functional information improves the biological interpretation of the observed cell response after virus infection. Although cytoskeletal proteins, mainly keratins, were overrepresented in both cases, the keratin cluster was connected to different nodes in the two solutions indicating that functional information influences the resulting network. The keratin cluster in the functional score solution also contained keratin 1 (KRT1) as part of the keratin module, thereby improving the keratin cluster. Proteins of interest from the original study (Liu, Song et al. 2008) such as Prohibitin (PHB) and cis-trans-isomerase A (PPIA) were also present indicating no important information had been lost by applying functional information to the network edges. Notably, the subgraph based on functional scores provided a more defined pathway connecting keratins to the rest of the proteins (over RAF1). The linking kinase RAF1 plays a major role under stress conditions and the phosphorylation-dependent disruption of RAF1 interaction with the keratin protein 18 (KRT18) has been previously described during stress response (Ku, Fu et al. 2004). The Ras GTP-ase (HRAS) found in the solution with functional scores activates RAF1 (Avruch, Khokhlatchev et al. 2001) in the signaling cascade and contributes to understanding how activation of RAF1 might have been triggered under stress conditions.

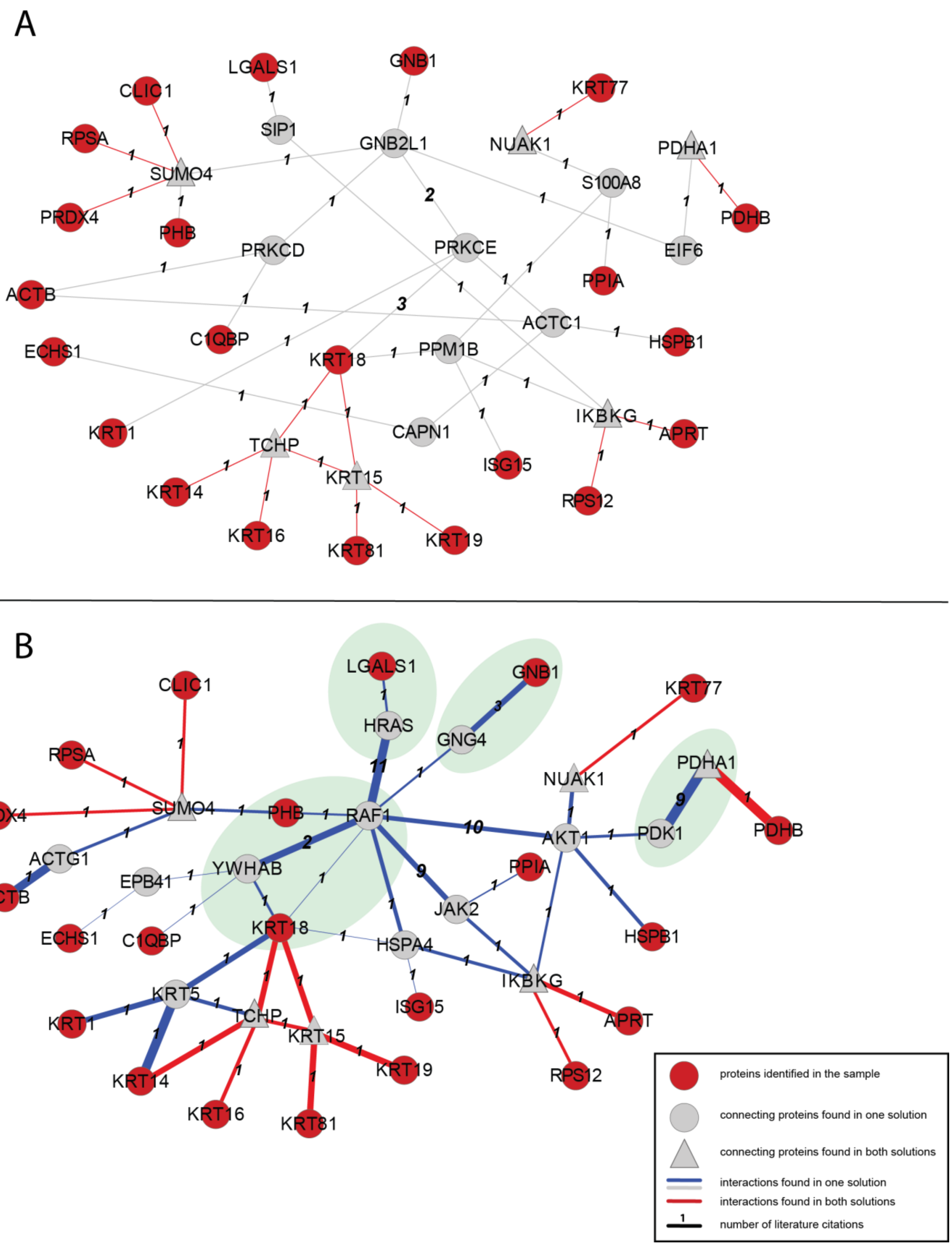


Figure 27. Comparison between networks from H9N2 virus infected human gastric cells extracted with (A) and without (B) functional interaction scores.

All 22 identified proteins were kept in both solutions and depicted as red circles. Linking proteins from the interactome are presented in grey. Triangles were used for representing linking proteins found in both solutions. Interactions contained in both solutions are marked in red, while interactions present only in the solution with functional scores are shown in blue. The edge width is proportional to the functional score of the interaction and the number of its source publications is marked on top. Thicker edges have a higher biological relevance and they are supported by a higher number of literature sources. We selected substructures based on the functional similarity of the module edges and their deviation from the module without edge scores (highlighted in blue). Results with induced GO semantic similarity show highly clustered functionally related proteins (Keratin cluster). RAF1 is a central player in this network (B), closely connected to the keratin cluters and the interaction between RAF1, YWHAB (14-3-3 protein) and KRT18 has a well-described role during cell stress.

To examine the differences between the two networks in respect to their interaction scores and functional similarity, we further considered the functional interaction scores of the network where they were initially disregarded (Figure 27A) and performed a Wilcoxon rank sum test to calculate whether there is a significant difference between the score sets of the two networks. Results are presented in Figure 28. With a p-value of 0.002 (median with scores: -1.9, without scores: -2.42; mean with scores: -1.73, without scores: -2.25) the solution with functional interaction scores showed a higher mean similarity score, thus, more of the interaction pairs in this solution were having a high functional similarity, as expected after introducing edge scores to the module detecting algorithm. Detailed statistics on the edge scores of both solutions can be found in Supplemental Tables 6 and 7.

This example demonstrates that functional interaction scores help to include interaction partners with a high functional similarity, which explain the complex biological effects on cell signaling after viral stimulation. Thus, the module extraction algorithm with interaction scores improves the functional relevance of the obtained clusters and focuses the resulting network on modules and paths containing substantial biological information needed for interpreting cell response. Proteins initially missing in the measured sample, but crucial for signal transduction can be identified using our method (e.g. RAF1).

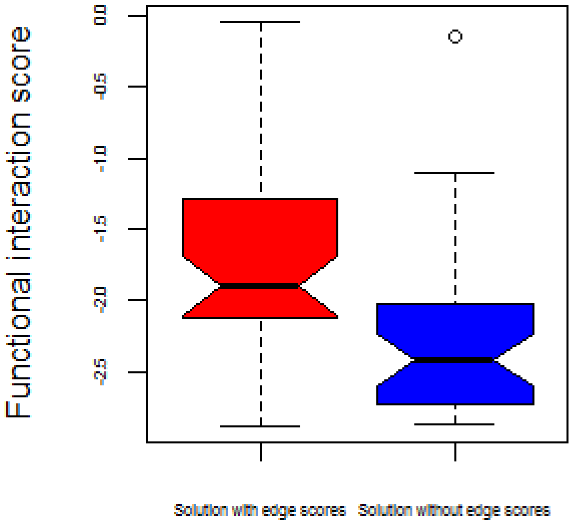


Figure 28. Boxplot of functional interactions scores

Functional interaction scores from the two network solutions were tested using a Wilcoxon rang test score. The solution with edges showed a higher median of edge scores (-1.9) when compared to the solution without scores (-2.420272). The mean in the first group (-1.73) was also higher than the mean of the second group (-2.25). The p-value after correction was 0.002. Therefore, the solution including edge scores had significantly higher functional scores chosen by the algorithm than the randomly chosen edges of the second solution.

5.3.3 Characterization of cell specific modules: T-cells

The large size of networks from proteomic analyses makes it difficult to search for a particular smaller functional module inside the resulting network. Therefore, we introduce a method for decomposing the network into smaller functional modules using the option for extracting suboptimal solutions contained in the module detection algorithm (Dittrich, Klau et al. 2008). The resulting network of T-cells from a blood constituents proteomi study (Haudek, Slany et al. 2009) was used as example.

From altogether 970 T-cell proteins, 861 were mapped to the human interactome. The approach was first used without considering functional scores. The proteins in the T-cell sample were given a constraint to ensure they are all present in the resulting network (as detailed in Materials and Methods chapter 4.9.6). The result using all T-cell proteins without applying interaction scores yielded a network of 1026 nodes and 3004 edges (Figure 29A). Naturally, a close look on single protein modules is not possible, therefore to reduce the size of relevant modules, we used an approach searching for the top five non-overlapping modules of a given size (in this case 50). In a second approach, the algorithm was performed again using functional interactions scores for extraction of the five size 50 modules. Finally, suboptimal solutions from both analyses (with and without interaction scores) were tested for functional enrichment.

In the suboptimal solutions without the use of functional interaction scores the p-values were very high as expected. In the cases where there was significant enrichment (p-value <= 0.05), biological process terms were broad and not specific to T-cell signaling. Thus, the module detection algorithm without interaction scores gives a general overview of the cells signaling and the extraction of characteristic modules is not possible. A more detailed summary of all enriched GO terms and their p-values can be found in Supplemental Table 8.

In the suboptimal solutions with functional interaction scores the optimal solution contained functional modules with high interaction scores, and the resulting functional modules represent more general biological processes such as ubiquitin/proteasome regulation (regulation of ubiquitin-protein ligase activity: 2.39×10^{-6}) and mRNA processing ($1.12 \text{ x } 10^{-5}$) (Figure 29B). The first suboptimal solution gave a result typical for T-cells: the terms T-cell differentiation (2.73×10^{-4}) and T-cell receptor signaling (2.73×10^{-4}) were enriched when compared to the set of all T-cell proteins (Figure 29C). Further enriched terms included signaling (3.81×10^{-5}), intracellular protein kinase cascade (1.09×10^{-5}) and signal transmission via phosphorylation event (1.09×10^{-5}). The second suboptimal solution contained enrichment of the following terms: cell movement, actin cytoskeleton organization (4.33×10^{-6}), leukocyte cell-cell adhesion (1.29×10^{-4}) and regulation of actin cytoskeleton organization (4.33×10^{-6}) (Figure 29D). Thus, the module detecting algorithm has managed to extract functionally similar modules from the network. A more detailed summary of all enriched GO terms and their p-values can be found in the Supplementary Table 9.

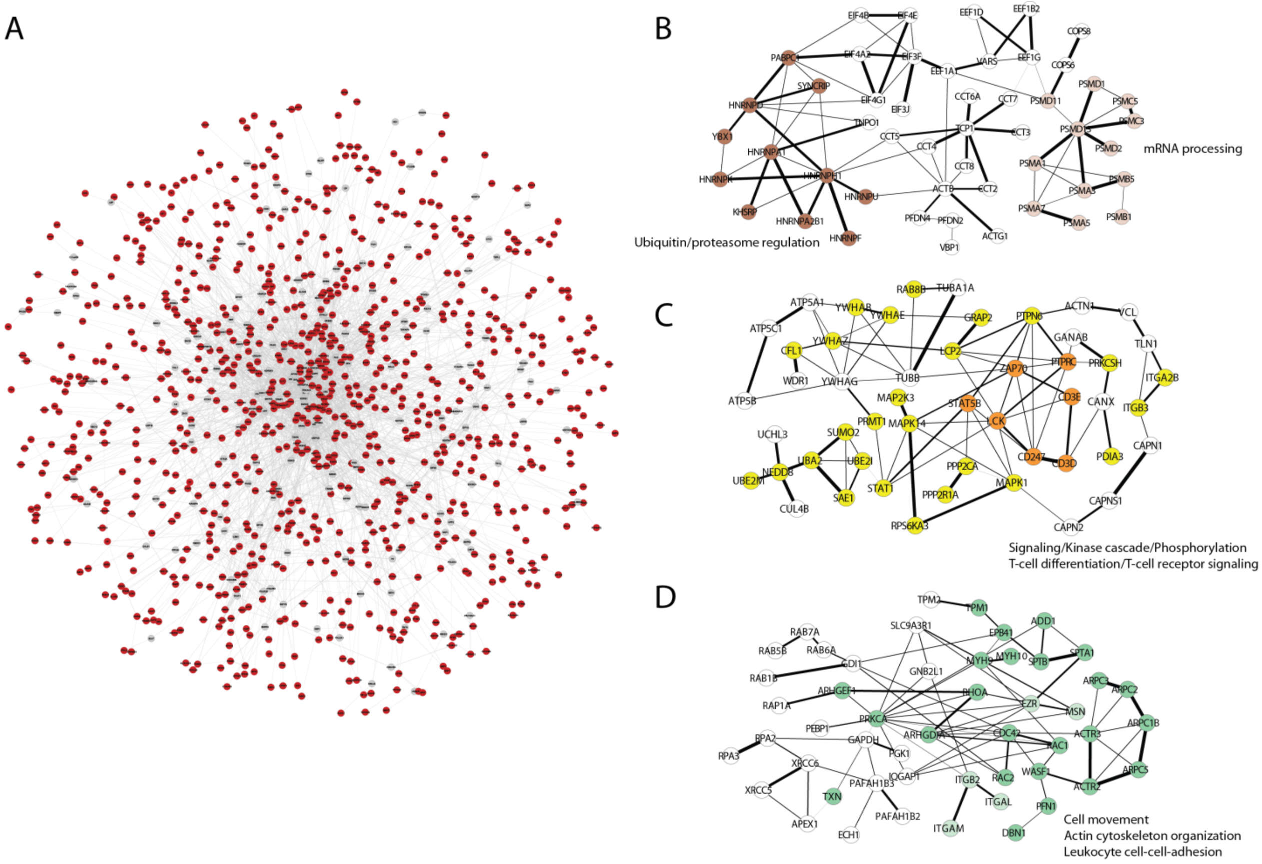


Figure 29. Functional modules of T-cells

A, Data was gathered from blood proteomics analysis of T-cells consisting of 972 proteins, from which 861 were mapped to the human interactome. Analysis without the use of functional interaction scores revealed a network of 1026 proteins tightly connected to each other through linking proteins from the interactome (proteins from the sample are depicted in red and linking proteins in the resulting network in grey). To decompose the network into non-overlapping functional modules, the algorithm was applied enhanced with functional interaction scores to obtain optimal (B) and suboptimal solutions (C, D) of 50 proteins. The thickness of edges represents the interaction score. To investigate the characteristic functional profile of T-cells, GO enrichment analysis against all T-cell proteins was performed on the resulting networks. B, Results revealed significant enrichment for general terms associated with ubiquitin regulation (brown) and mRNA processing (light red) in the optimal solution. C, Terms characteristic for the T-cell type such as signaling cascades, protein phosphorylation (yellow), T-cell differentiation and T-cell receptor signaling (orange) were enriched in the first suboptimal solution. D, Regulation of cell movement, actin cytoskeleton organization (green) and leukocyte cell-cell adhesion (light green) were enriched in the second suboptimal solution. Thus, the network could be decomposed in its most characteristic functional modules.

5.4 Quantitative phosphoproteomic analysis

Quantitative phosphoproteomics is a newly evolving technology in the field of mass spectrometry (Macek, Mann et al. 2008; Ozlu, Akten et al. 2010). The advantages analyzing phosphorylation sites and measuring quantitative changes in phosphorylation over time has become indispensable for the investigation of complicated whole cell analysis under various condition. As phosphorylation is one of the main signaling mechanisms in human cells, measuring phosphorylation changes precisely in an automatic way means sensitive measuring of cell responses to various types of stimuli. Furthermore, understanding network signaling through phosphorylation can lead to the identification of novel biological markers and the development of new pharmacological therapies (Erler and Linding). The link between conserved phosphorylation sites and their impact in multiple diseases has also been indicated in a recent study by Tan et al (Tan, Bodenmiller et al. 2009). Many recently published datasets investigate phosphorylation changes on a global scale in important cell processes such as mitosis (Nousiainen, Silljé et al. 2006; Malik, Lenobel et al. 2009; Olsen, Vermeulen et al. 2010), signaling pathways (Choudhary and Mann 2010) and cellular sub-compartments (Boja, Phillips et al. 2009). These studies identified a large amount of proteins with changed phosphorylation but their interpretation in a network signaling context remains a challenge.

While phosphorylation changes in Hela cells and embryonic stem cells have progressed in recent years, there has been sparse information available on quantitative phosphoproteomic changes in human platelets. This was finally achieved in collaboration with groups from our consortium (**S**ystems biology of PGI2 and **A**DP P2Y12 **R**eceptor signaling: **SARA**). Tight collaboration with the clinical biochemistry group of Prof. U. Walter and mass spectrometry analysis of Prof. A. Sickmann supplied a number of phosphorylations quantitatively measured in human platelets. I performed a detailed analysis of the platelet signaling changes by applying our module detection algorithm (Dittrich, Klau et al. 2008). In a further step, this algorithm has been extended for the investigation of large-scale quantitative phosphoproteome data using a human embryonic dataset (Rigbolt, Prokhorova et al. 2011). I interpreted the results biologically and optimized the algorithm for better performance.

5.4.1 Quantitative phosphoproteomics in stimulated human platelets

For the following analysis, platelets were extracted from blood of patients and analyzed using mass spectrometry label-free quantification methods (proteomics lab of Prof. Albert Sickmann). Platelets were stimulated in time series by Iloprost (10s, 30s, 60s, 120s) and ADP (10s, 30s, 60s). Iloprost is a Prostacyclin analogue, which binds to the platelet prostaglandin I2 (prostacyclin) receptor (PTGIR) and causes platelet inhibition by increase of cAMP and PKA activation (Broos, Feys et al. 2011). In contrast, ADP binds to the platelet P2RY12 receptor and triggers downstream activation of platelets by inhibiting PKA (Jin, Quinton et al. 2002)(see 3.2.1 ADP signaling).

At first, I added all peptides and source information to tables in MySQL for further mapping. Then, all identified peptides were mapped to their protein sequences in HPRD using a Perl script. Thus, phosphorylation sites were correctly identified with their position in the HPRD protein sequence. Data from the source quantitative measurements were combined with the phosphorylation mappings and kinase-substrate information was subsequently extracted from *PlateletWeb* based on HPRD information. Finally, each phosphorylation site had an associated protein sequence, kinase data, and quantitative measurements after stimulation with ADP and Iloprost.

The total number of distinct identified phosphorylation sites was 596 after mapping to the human interactome using HPRD sequence information. Around 70% of these sites (421) were already known from experiments in other cell types, but 148 sites had no source information available. Only 40 sites were associated with a kinase (37 kinases), indicating the scarcity of available regulation data. As Iloprost information is connected with PKA activation, we tested for present PKA sites. There were 10 PKA sites found, which were already described in literature (HPRD or PhosphoSite) and the number of classical PKA sites found after motif search was 66. Out of these, only six were confirmed in literature as PKA targets, 49 were already measured in other studies and 17 were novel phosphorylation sites.

Based on this basis information a module detection algorithm was used to identify functional modules in the given dataset (Dittrich, Klau et al. 2008). This algorithm assigns positive and negative values to the interacting nodes based on the priority and quantitative measurements given to the proteins. Using the human PPI network as a backbone, it searches for the maximum-scoring subnetwork in a given list of proteins (Materials and Methods). Proteins from the ADP model were given a high constant value of +10 to ensure that they are considered in the solution. All proteins from the quantitative phosphoproteomics dataset were assigned the sum of the log2 values of all time points (sumlog2=(log2(1)*log2(1)+log2(2)*log2(2)). The log2 was calculated for the ratio between each time point and the control. The concept of considering the sum of all log values over all time points is to identify proteins with maximally changed phosphorylation sites. By multiplying the values we made sure that results remained independent of whether phosphorylation increased or decreased over time. For further investigation, the algorithm was also calculated for all separate time points (data not shown).

Proteins extracted from the ADP model were given a very high score of +10 (Iloprost) and +20 (ADP) to ensure their presence in the final subnetwork. All measured proteins were assigned the sumlog2 value of the most changing site over time. The rest of the interactome proteins were split into two groups: platelet proteins were given a constant value of the negative average of the sumlog2 value of all proteins measured during Iloprost stimulation. The non-platelet proteins were assigned a value of -1 as they have the least importance in the conducted analysis. The results of the analysis are presented in Figure 30 for Iloprost stimulation and Figure 31 for ADP stimulation.

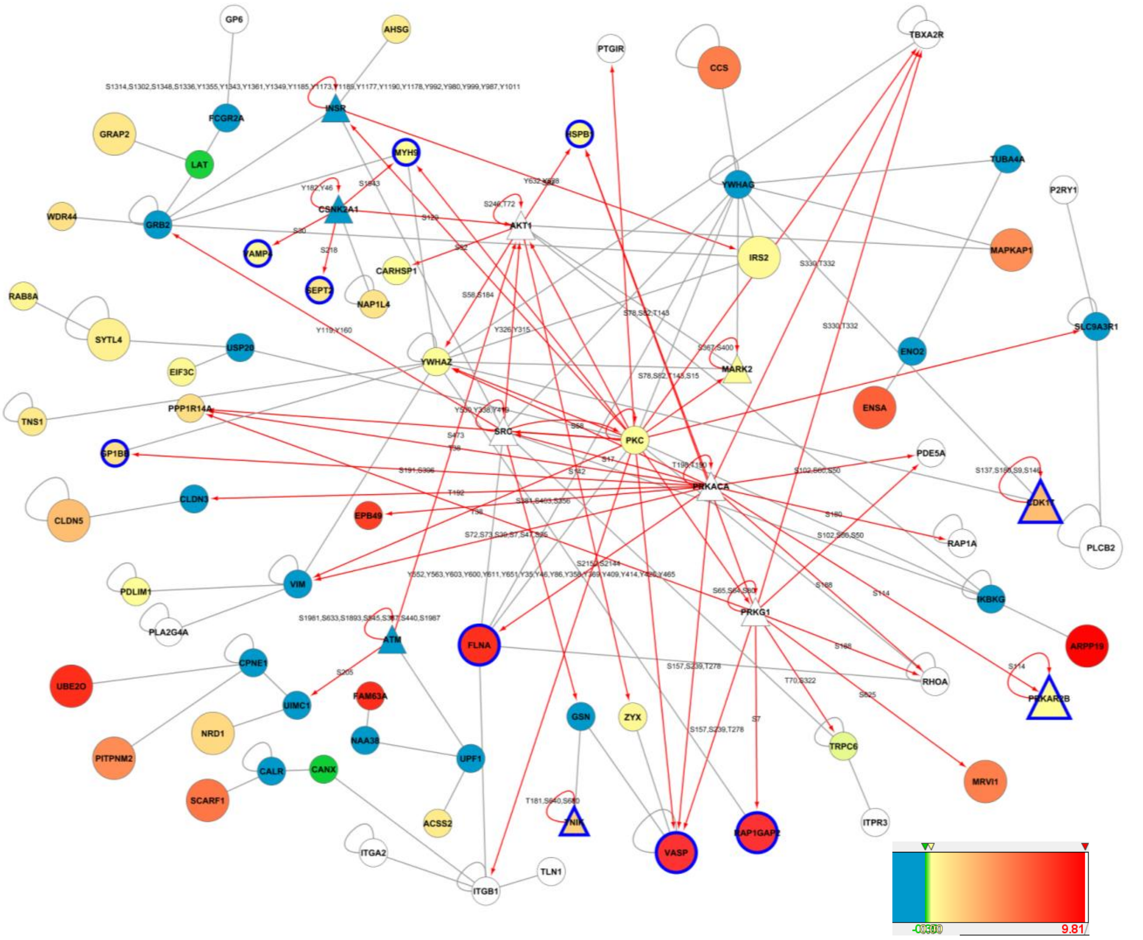


Figure 30. Phosphorylation response module after platelet Iloprost stimulation

The platelet response module after Iloprost stimulation. All four time points are taken and the sum of the logarithmic ratios of each time point against the control is calculated as representative of the overall phosphorylation change. Kinases are presented as triangles, proteins as circles and interactions are depicted with grey lines. Red arrows represent phosphorylation events. The scores are shown in a range from green to red with red being the highest score. The green proteins have a negative value because the original sumlog values are transformed into a lower range of numbers as the algorithm performs better with a distribution of values around the 0 point. A value of **0.5** was additionally substracted from the original summed logarithmic value.

Further coloring includes blue for linking proteins from the interactome and white for ADP model proteins. A big node size represents proteins with classical PKA sites, which were measured in the experiment and a blue circle around the node is drawn when the phosphorylation site is matching a known phosphosite targeted by the depicted kinase in literature experiments (extracted from *PlateletWeb*).

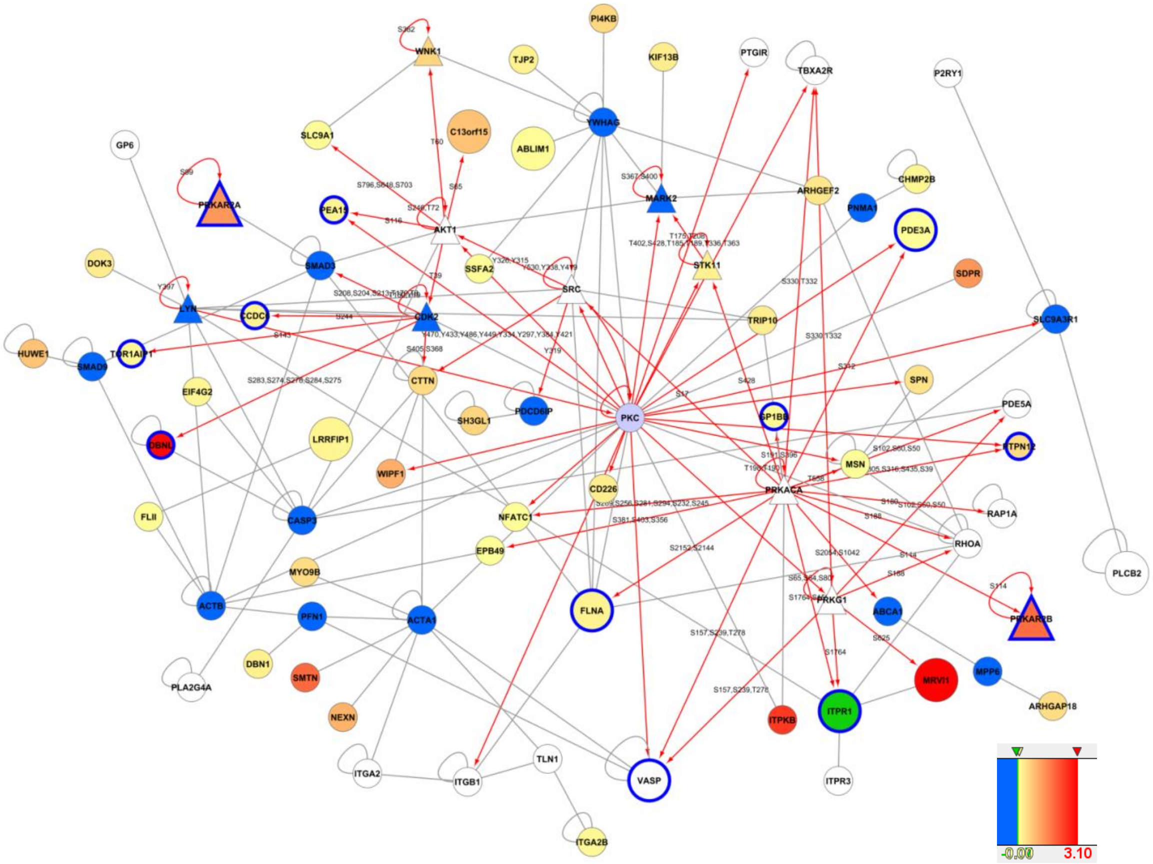


Figure 31. Phosphorylation response module after platelet ADP stimulation

The platelet response module after ADP stimulation. All four time points are taken and the sum of the logarithmic ratios of each time point against the control is calculated as representative of the overall phosphorylation change. Kinases are presented as triangles, proteins as circles and interactions are depicted with grey lines. Red arrows represent phosphorylation events. The scores are shown in a range from green to red with red being the highest score. The green proteins have a negative value because the original sumlog values are transformed into a lower range of numbers, because the algorithm performs better with a distribution of values around the 0 point. A value of **0.2** was additionally substracted from the original summed logarithmic value.

Further coloring includes blue for linking proteins from the interactome and white for ADP model proteins. A big node size represents proteins with classical PKA sites, which were measured in the experiment and a blue circle around the node is drawn when the phosphorylation site is matching a known phosphosite targeted by the depicted kinase in literature experiments (extracted from *PlateletWeb*).

Ilorpost stimulation causes an overall increase in phosphorylation. The proteins with strongest changes in phosphorylation, Filamin A (FLNA), VASP and RAP1GAP2 (a protein activating the small GTPase Rap1 in platelets), are all also phosphorylated on the known PKA site, which is consistent with the observation after Iloprost stimulation. Filamin A and VASP are both inhibited by PKA phosphorylation and they are hindered in facilitating cytoskeleton organization and platelet activation. The ubiquitin-conjugating enzyme E2O (UBE20) shows a significant change in phosphorylation. The changes in the cAMP-regulated phosphoprotein (ARPP19) are also indicating PKA activity and are to be expected in this experimental setting. Dematin, or EPB49, is an actin-bundling protein originally identified in the erythroid membrane skeleton. Its actin-bundling activity is abolished upon phosphorylation by PKA and is restored after dephosphorylation, which explains the changed phosphorylation after Iloprost stimulation (Communi, Vanweyenberg et al. 1997). The changes of phosphorylation of endosulfine alpha (ENSA) which belongs to the cAMP-regulated phosphoprotein (ARPP) family as well, can be also explained by the activation of PKA.

In the ADP resulting network, two proteins show high change in phosphorylation: murine retrovirus integration site 1 homolog (MRVI1) and inositol 1,4,5-trisphosphate 3-kinase B (ITPKB).

Analysis of the networks revealed an overall increase of phosphorylation after Iloprost stimulation. The measured enrichment of PKA sites was also consistent with the activation of PKA during Iloprost stimulation. The ADP results on the other hand indicated downregulation of phosphorylation during ADP stimulation and it is unclear, whether this finding is based on real facts or arises as a possible artifact from technical problems. Overall, changes in phosphorylation in both directions (up or down regulation) can be visualized in a network context by choosing the optimal solution, which links most of the measured proteins with ADP proteins from the model. Thus, hypothesis about putative signaling pathways and regulation arise which can then lead to further lab experiments.

Unfortunately, the provided data was associated with some technical difficulties in the mass spectrometry detection methods. Therefore, this data could not be further used, but the analysis still represents a successful starting point in future quantitative phosphoproteomics investigation of human platelets.

5.4.1 Identification of functional modules in quantitative phosphoproteomic data from human embryonic stem cells (hESCs)

The increasing amount of proteomic and phosphoproteomic data has motivated the development of new approaches in integrative network analysis. Differences in protein abundance and phosphorylation levels represent changes in cellular signaling which are fully comprehensible only in a systems biological manner. Embryonic stem cells are important in research due to their ability to differentiate into various tissues and organs (Thomson, Itskovitz-Eldor et al. 1998). This ability is controlled by a number of transcription factors and fascinates researchers by the intricate regulation during maintenance of pluripotency and lineage specification (Thomson, Itskovitz-Eldor et al. 1998). Phosphorylation plays a crucial role during differentiation and Rigbolt et al performed one of the most comprehensive phosphoproteomic studies in embryonic stem cells to date using stable isotope labeling by amino acids in cell culture (SILAC) (Rigbolt, Prokhorova et al. 2011). In the used study, phosphorylation changes were measured in human embryonic stem cells at four time points after stimulation with non-controlled medium (NCM) (see Materials and Methods). In this medium, factors needed for cell differentiation were removed. The study identified 6521 phosphorylated proteins and overall 23522 phosphorylation sites. The identified phosphorylation sites in the dataset were filtered into class 1 sites, which by definition were identified with a probability of at least 0.75. Considering only these sites (15,004) and additionally missing some due to mapping and converging IPI identifiers with Entrez gene identifiers, we identified 13842 distinct phosphopeptides with 156 kinases acting on these phosphosites. From all measured phosphosites, 641 peptides and 363 proteins were associated with at least one kinase. The SILAC ratios were calculated in order to obtain the differential phosphorylation between the treated and the controlled cells (Figure 32). Then, the dataset proteins were combined with the kinases targeting these sites from literature to create a phosphoproteome network with 343 nodes, which was visualized in Cytoscape (Shannon, Markiel et al. 2003). There were 262 nodes coming from the dataset, of which 39 were kinases. From the human interactome, we extracted additional 81 kinases targeting the measured sites. The network was densely connected with 285 nodes contained in the largest connected component (84 % of the whole network).

We extended our investigation of functional modules in the human proteome by including quantitative phosphoproteomics data and weighting nodes of the network according to the kinase and substrate information of each protein. Site specificity was achieved by zooming down to single sites representing each protein, while kinase-substrate relationships were extracted from the *PlateletWeb* repository (Boyanova, Nilla et al. 2012). Thus, all depicted phosphorylations have been identified previously in human cells. The module detection algorithm (Dittrich, Klau et al. 2008) was performed according to the scheme in Materials and Methods but in this case the node scores of all identified proteins represented the ratio between the phosphorylation measured in each time point against the zero time point. Only the maximally changing phosphorylation site over time was taken into consideration. Thus, the network represents single phosphorylation sites measured in each protein (node) and the absolute log2 transformation of the ratio for each time point versus the control time point of the maximally changing phosphosites was assigned to the nodes of the network. Kinases were additionally given a higher score, to ensure that the algorithm considered them with higher priority than the rest of the proteins. Kinases extracted from literature, which phosphorylated the measured site, obtained a constant kinase score of 0, while kinases measured in the experiment kept their measured values, which were also logarithmically transformed to build a score as explained above (by definition their value was higher than the value for kinases from the phosphoproteome (0)). As the algorithm searches for the maximum scoring subgraph, kinases are more likely to appear in the final solution and thus the network is by definition more concentrated on the regulatory signal flow.

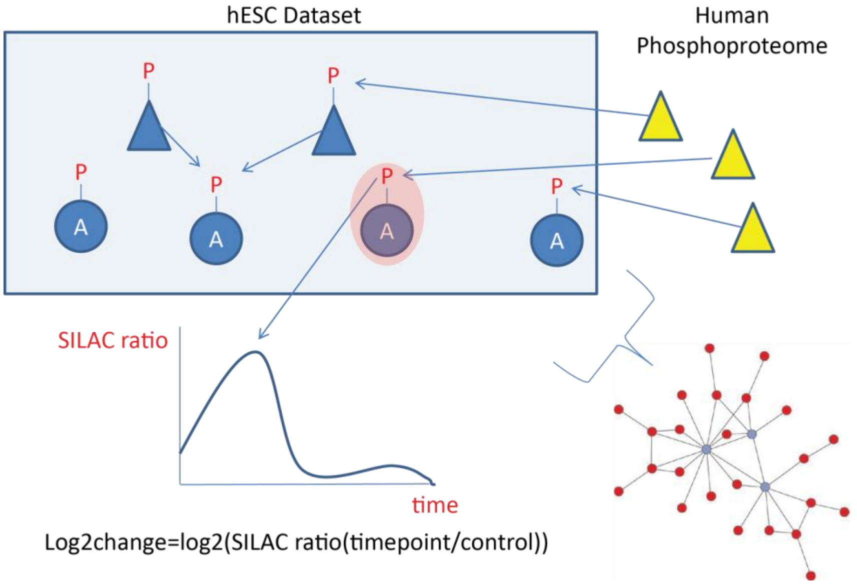


Figure 32. Experimental design of the phosphoproteomics analysis

The experimental design consisted in combining information from the hESC dataset with human phosphoproteome data for kinases, acting on the measured sites. The SILAC ratio between each time point and the control time point was considered as measurement of the phosphorylation change. Subsequently, time-specific response modules of phosphorylation signaling during hESCs differentiation were extracted using the module detection algorithm (Dittrich, Klau et al. 2008).

Results of the module detection algorithm were presented as follows: The networks were superimposed as a union of all nodes extracted for each time point, thus keeping the network topology constant so that the changes of phosphorylation can be traced over all time points on all proteins (Figure 33). Nodes represent the site with maximum changing phosphorylation over all four time points. The resulting networks were visualized with the node color representing the original ratios of all phosphorylation sites in a range from blue (phosphorylation decrease) to red (phosphorylation increase). The network represents site-specific events (the node depicts the phosphorylation ratio of the maximum changing site against the control), therefore the changing sites can also be investigated closely. A number of proteins increased their phosphorylation within the four time points. Kinases from the experimental dataset such as MAPK1, MAPK3 and JUN increase their phosphorylation at particular sites, while other kinases from the dataset such as CDK2 decrease their phosphorylation (details in Table 3). These changes are particularly interesting, because these kinases have been identified in the measured sample and may be responsible for the observed effects on other proteins. Listing the kinases phosphorylating the maximally changing phosphorylation sites of key proteins (Table 3) reveals that MAPK1, MAPK3 and CDK2 phosphorylate a number of substrates at the exact positions identified in the sample. MAPK1 and MAPK3 phosphorylation is increasing towards 24h, while JUN kinase phosphorylation has a significant peak at 6h. The phosphorylation site Ser^{63} on JUN has already been indicated to change during hESC differentiation (Van Hoof, Dormeyer et al. 2010), in our study we reveal a possible regulation of this event (by kinases: MAPK9, MAPK8, VRK1, MAPK15, PLK3).

A number of proteins have an increase in their phosphorylation during NCM stimulation, such as Vimentin. This protein is a member of the intermediate filament family and it is responsible for maintaining cell shape, integrity of the cytoplasm and stabilizing cytoskeletal interactions (Goldman, Khuon et al. 1996). Vimentin phosphorylation at Ser^{10} increases after 24h and this phosphorylation, carried out by PKC alpha, is responsible for disassembly of the vimentin filament structure (Ando, Tanabe et al. 1989), which might have implications during hESC differentiation.

CDK1 and CDK2 have been identified in a hESC differential phosphorylation analysis by van Hoof et al (Van Hoof, Dormeyer et al. 2010) and they were indicated as central in controlling self-renewal and lineage specification. The same study also identified the Vimentin phosphorylation, but not on the site Ser^{10}, which might be a new site with importance for lineage specification in embryonic stem cells.

The transcription factor ETV6, also known as Leukemia-Related Transcription Factor TEL, is phosphorylated by MAPK3 (ERK1) at Ser^{213} which inhibits the action of the transcription factor (Maki, Arai et al. 2004). A previous study already determined that ETV 6 is required for hematopoietic stem cell maintenance (Akala and Clarke 2006), which indicates that the inhibition of the transcription factor through phosphorylation may be preventing stem cell renewal in the hESC cells. However, this has to be experimentally validated.

Another protein with upregulated phosphorylation is the Na+/H+ antiporter SLC9A1, with maximum phosphorylation at 24 h after triggered differentiation.

Some proteins show variability in phosphorylation changes, at first increasing and then decreasing their phosphorylation state. For example the SNW1 protein which is phosphorylated by CDK2 becomes phosphorylated at first and the phosphorylation of the site decreases with decrease of CDK2 phosphorylation, indicating CDK2 inhibition might occur due to initial phosphorylation decrease of its site Tyr^{15}. This could be confirmed by analyzing the Tyr^{15} phosphorylation in detail. Previous studies indicate that this is an inhibitory site phosphorylated in higher eukaryotes during the cell cycle and the phosphorylation is carried out by the kinase WEE1. Their results suggest that the activity of WEE1 is regulated by phosphorylation and proteolytic degradation, and that WEE1 plays a role in inhibiting mitosis before M phase by phosphorylating cyclin B1-Cdc2 (Watanabe, Broome et al. 1995).

Our investigation of the embryonic stem cell phosphoproteome concluded that the site-specific phosphorylation network, complemented by kinase regulation from the proteome study and the human phosphoproteome, gives a systemic view on the changes observed in the cellular system. The signal flow can be followed precisely, based on information from experimental data obtained in literature and important changes during hESC differentiation can be studied in a site-specific manner. Thus, this approach enhances considerably the initial analysis of top proteins found in the proteomic list.

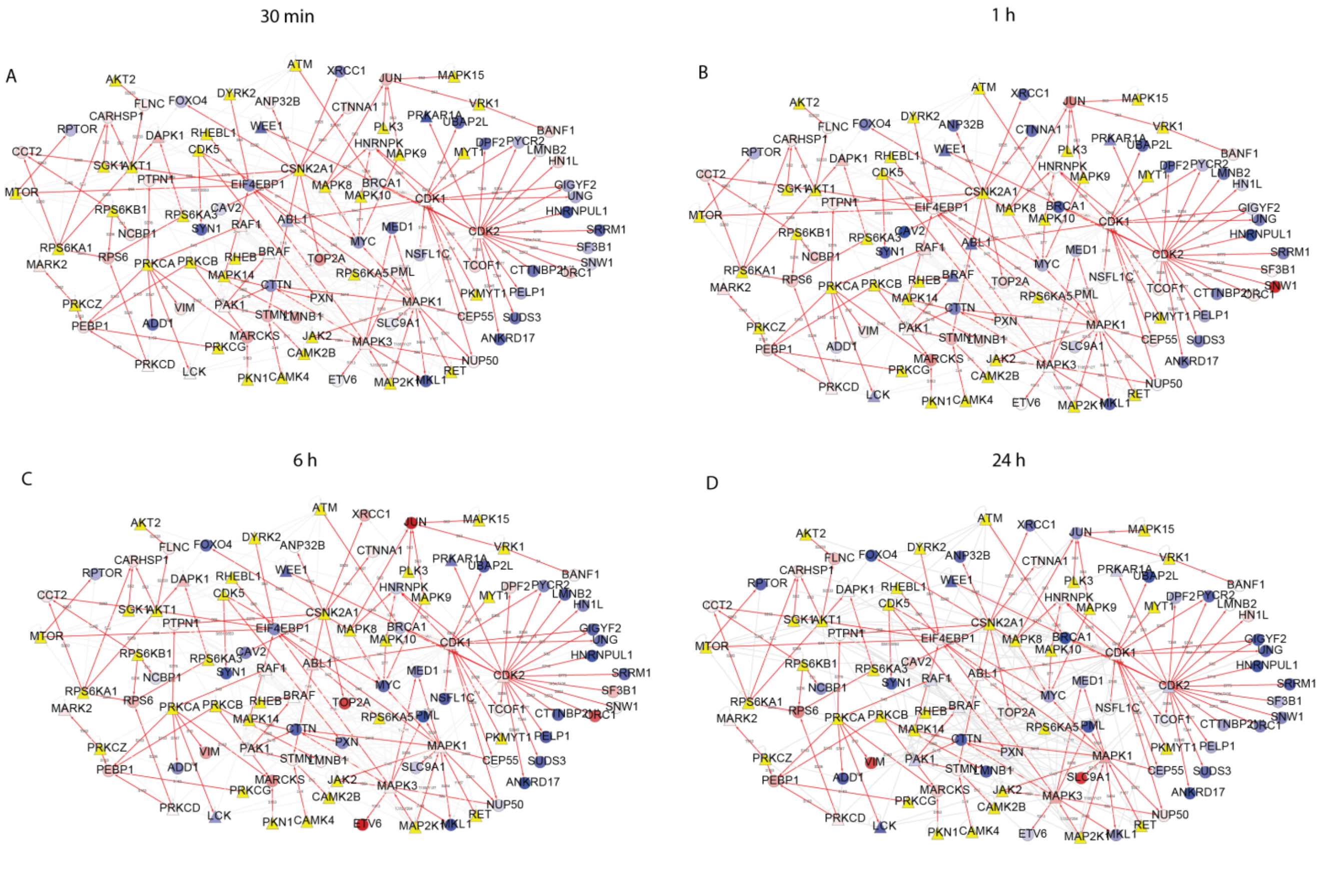


Figure 33. ESC stimulation response modules.

The network represents the resulting maximum-scoring subnetworks of the module detecting algorithm (Dittrich, Klau et al. 2008) after (A) 30 min, (B) 1h, (C) 6h and (D) 24h of stimulation for differentiation. The networks are superimposed as a union of all single modules extracted for each time point, thus keeping the network topology constant so that the changes of phosphorylation can be traced over all time points. Nodes represent the site with maximum changing phosphorylation over all four time points. The node color depicts the absolute logarithmic ratio (log2) of the phosphorylation value between the measured time point against the control (time point 0). There are proteins with increased or decreased phosphorylation, as well as proteins with perturbations over all four time points (ORC1, SNW1, both increase at first and decrease at time point 24h).

Protein Name	***Description***	***Phosphosite***	***Up/Downregulation***	***Kinase***	***Measured in dataset***
VIM	Vimentin	Ser^{10}	↑ (max. 24h)	PRKCA	no
ETV6	ets variant 6	Ser^{213}	↑ (max. 6h)	MAPK3	yes
TOP2A	topoisomerase (DNA) II alpha 170kDa	Ser^{1377}	↑ (max. 6h)	CSNK2A1	no
JUN	jun proto-oncogene	Ser^{63}	↑ (max. 6h)	MAPK9,MAPK8, VRK1,MAPK15,PLK3	no
MAPK1	mitogen-activated protein kinase 1	Tyr^{187}	↑ (max. 24h)	JAK2,MAP2K1,RET	no
MAPK3	mitogen-activated protein kinase 3	Tyr^{204}	↑ (max. 24h)	MAP2K1	no
SLC9A1	solute carrier family 9 (sodium/hydrogen exchanger) member 1	Ser^{785}	↑ (max. 24h)	MAPK1	yes
FOXO4	forkhead box O4	Ser^{230}	↓	CXNK2A1	no
RPTOR	regulatory associated protein of MTOR, complex 1	Ser^{863}	↓	MTOR	no
LCK	lymphocyte-specific protein tyrosine kinase	Ser^{42}	↓	PRKCA	no
PML	promyelocytic leukemia	Ser^{530}	↓	MAPK1, MAPK3	yes
ANKRD17	ankyrin repeat domain 17	Ser^{2045}	↓	CDK2	yes
GIGYF2	GRB10 interacting GYF protein 2	Ser^{30}	↓	CDK1,CDK2	yes
BRCA1	breast cancer 1, early onset	Ser^{1497}	↓	CDK1,CDK2, ATM	yes
CTTN	cortactin	Ser^{418}	↓	PAK1	yes
ANP32B	acidic (leucine-rich) nuclear phosphoprotein 32 family, member B	Thr^{244}	↓	CSNK2A1	no
NCBP1	nuclear cap binding protein subunit 1, 80kDa	Thr^{21}	↓	RPS6KB1	no
ADD1	adducin 1 (alpha)	Ser^{747}	↓	PKRCA	no
NSFL1C	NSFL1 (p97) cofactor (p47)	Ser^{140}	↓	CDK1	yes
CDK2	cyclin-dependent kinase 2	Tyr^{15}	↓	WEE1	yes
LMNB1	lamin B1	Ser^{405}	↓	PKRCB	no
ORC1	origin recognition complex, subunit 1	Ser^{273}	6h↑;24h↓	CDK2	yes
SNW1	SNW domain containing 1	Ser^{224}	1h↑; 24h↓	CDK2	yes

Table 3. Differentially phosphorylated sites among all four time points of differentiation stimulation in human ESCs

The table represents proteins with changed phosphorylation sites, included in the solution of the module detection algorithm. The exact sites are depicted along with an arrow showing up- or downregulation of the site. The last column on the right holds information whether the kinase responsible for phosphorylating the site was also measured in the dataset or originates from the human phosphoproteome information extracted from literature.

6 Discussion

This thesis established approaches for analysis of the cell proteome using integrated network analysis and proteomics data. The developed knowledge base *PlateletWeb* integrates data from platelet protein detection studies, literature knowledge on interactions, phosphorylations and protein function coupled with drug and disease associations (Boyanova, Nilla et al. 2012). The main aim of this website was to present platelet researchers with a tool for systems biological analysis of single platelet proteins or a list of platelet proteins of particular interest. The assembly of the *Plateletweb* knowledge base allowed global investigation of the platelet proteome revealing that platelet proteins were mainly extracted from whole platelet lysates in proteomics studies. A closer look on phosphorylation and dephosphorylation events suggested that only a few phosphatases were responsible for counteracting kinase function. An enrichment analysis for platelet drug targets indicated that proteins with a higher connectivity in the platelet proteome are more often targeted by drugs. Closely investigating this finding, we confirmed that mainly experimental drugs targeting proteins with kinase activity were responsible for this effect.

In a second approach, various signaling pathways of the platelet were analyzed using integrated networks with interaction and phosphorylation information extracted from *PlateletWeb*. ADP signaling and SH2 domain binding proteins were analyzed in a network context for collaboration studies and kinase predictions for experimentally validated phosphorylation sites in platelets gave rise to a new hypothesis of integrin inside out signaling based on DOK1 inhibition and phosphorylation.

Using the PPI interactome as backbone, an approach for functional module detection in cells was validated based on functional interaction scores of the interactome edges. Functional characterization of T-cells and analysis of small and medium-sized proteomic datasets illustrated the useful applications of this method.

In a final study, my thesis presented an integrative approach to quantitative phophoproteomic analysis starting with a tutorial using platelet-specific phosphorylation sites measured by a collaboration group after platelet activation/inhibition, and further covering a detailed phosphosite network analysis in embryonic stem cell dataset of site-specific phosphorylations after trigged differentiation. Using both datasets, I showed an improvement in identifying response functional modules after cell stimulation based on our algorithm (Dittrich, Klau et al. 2008).

In this chapter, I would like to present a discussion for all important findings in this thesis and underline their contribution to the fields of bioinformatics and integrated network analysis.

6.1 Unraveling the platelet proteome

The platelet proteome has been collected for the first time in a platelet-specific platform called *PlateletWeb*. This knowledge base gives opportunities to investigate platelet proteins in a network context, using integrated information from their modifications (phosphorylation and dephosphorylation), kinase and phosphatase regulation, drug and disease associations along with their functional and physical properties. The website allows the analysis of each protein individually or the investigation of a group of proteins with similar functions and characteristics using the advanced search options. Using integrated network analysis of the available information, researchers can generate hypothesis about new interesting phosphorylation sites based on the neighboring network. Experimental data from the lab can also be used directly to identify proteins of particular isoelectric point or molecular weight ranges. In a more advanced approach, physical properties can be coupled with functional annotations to limit a particular group of platelet proteins. Own lists of proteins can also be used and visualized in a network context using the subnetwork extraction option, thus rendering various opportunities for systems biological analysis in platelets based on experimentally validated, retractable and traceable information. Thus, a first insight on platelet signaling in the light of systems biology could be provided for the platelet community (Boyanova, Nilla et al. 2012).

By introducing kinase predictions hypotheses for the regulation of 81,8 % of the newly discovered phosphorylation sites could be generated enabling analysis of the signaling regulation in platelets under basal conditions. Kinase predictions may therefore prove very useful in phosphoproteomics research as many of the identified phosphorylation sites lack kinase association information at first. These predictions give indications on which kinases can be tested for activation and phosphorylation of the given site and thus save valuable time in experimental procedures. On the other hand, they are an important bioinformatical asset for the analysis of signaling transduction as network analysis is only achieved if proteins are connected to each other and the phosphorylating kinase is available. Kinase predictions can further extend our knowledge not only for platelet regulation but also for any other cell line, as there is not yet a comprehensive list of all phosphorylation regulations in human cells. When considering platelet phosphatases, dephosphorylation is achieved by fewer enzymes, which have a broader specificity.

The other important aspect of platelet function – transmembrane domain analysis – revealed the presence of transmembrane protein predictions, for which experimental validation is still missing. As the platelet membrane proteome presents challenges during proteomics analysis, predictions may add some important new insights on putative receptors and transmembrane proteins in the platelet. Transmembrane predictions have extended the information about platelet proteins and revealed new proteins previously not identified in membrane-specific studies. Therefore, the use of such predictions enriches the platelet database, as it allows an estimation of the number of transmembrane proteins in the platelet proteome. Further detailed analysis can be performed using this information in a network context or in functional experimental studies.

The proteome analysis of platelets continued with pathway enrichment analysis. The identification of endocytosis as the top significant pathway found in platelets is consistent with platelet involvement in inflammation and potentially in metastasis (Leslie 2010) and supported by experimental findings of a recent off-gel proteomics study analyzing pathways with platelet proteins (Krishnan, Gaspari et al. 2011). Endocytosis is the internalization of plasma membrane proteins, lipids, extra-cellular molecules, bacteria and viruses and actin cytoskeleton is required for these processes (Pelkmans and Helenius 2002). There are different mechanisms for endocytosis (clathrin-mediated, virus entry, caveolar-mediated endocytosis), from which the virus entry is an interesting topic connected to platelets. At first, it was assumed that platelets are only involved in hemostasis but later new functions have been revealed (Leslie 2010). There is a growing evidence of platelet interaction with pathogens. Various bacteria have shown capability of binding to platelets directly through cooperation between the αIIbβ3 and FcγIIa, for example (Ludwig, Schultz et al. 2004). In most cases the pathogen entry is facilitated by binding to a plasma protein, which then interacts with platelet receptors such as integrins (Kerrigan and Cox 2010).

Unexpected over-represented pathways related to pathological conditions include Alzheimer's disease pathway, mirroring the recently discovered role of platelet APP as biomarker for this disease (Borroni, Agosti et al. 2010). A correlation was detected between the platelet APP ratio and the severity of disease in an early stage of Alzheimer's disease (Padovani, Borroni et al. 2001). The overrepresentation of actin cytoskeleton is to be expected, as the cytoskeleton plays a crucial role in the shape change of platelets (Bearer, Prakash et al. 2002).

Functional aspects of the platelet proteome and analysis of overrepresented pathways are tightly connected with pharmacological therapies, targeting platelet proteins. To reveal the number of drug targets in platelets based on the assembled proteome and drug data we performed various analyses with platelet drug targets. Essential proteins are well known to be highly connected hubs in biological network (Jeong, Mason et al. 2001) and this has also been reported for drug targets (Yildirim, Goh et al. 2007). Whereas most of the previous studies analyzed a 'generic' human network comprising the set of interactions known at the time of study, we examined here the cell type specific network of anucleate human platelets. Similar as in 'generic' human PPI networks we also observe a higher connectivity of drug targets and disease-associated genes in the PPI network of platelets. Concerning the disease-associated genes this effect might be due to the influence of essential disease-associated genes, as previous studies suggested the correlation with network degree could not be observed after the removal of essential genes (Goh, Cusick et al. 2007). For platelet drug targets the separate analysis of experimental and approved drug targets clearly demonstrates that this association applies mainly to the experimental targets.

To obtain a deeper understanding in how regulation patterns of the network can be modulated via therapeutical approaches we analyzed the connectivity of drug targets in the signaling network of direct kinase-substrate relationships. For this we find a similar association with higher degree (i.e. number of substrates) for drug targeted kinases, whereas for disease- associated kinases no difference in average number of substrates could be detected. Interestingly, in a previous study site-specific phosphorylations of substrates have been investigated and phosphorylation hubs on the side of the substrates were found to include more disease-related genes (Tan, Bodenmiller et al. 2009). Obviously, a correlation with connectivity of disease related genes becomes apparent only on the side of the substrates but not on that of kinases.

The separate analysis of experimental and approved drugs shows that highly connected kinases are significantly more often drug targets of experimental drugs, suggesting that current pharmacological efforts mainly focus on kinases with a broad specificity, having a large number of substrates in the platelet network. Among the drug-target platelet kinases with more than 30 substrates over 90% are associated with experimental drugs and 50% are not yet targeted by approved drugs. Among these is protein kinase A (PKA), a major inhibitory kinase in platelets, which phosphorylates 105 platelet substrates and is targeted by 44 experimental drugs, most of which are kinase inhibitors such as phosphonoserine. Experimental drug targets developed for various applications can be an important new resource for platelet pharmacology, as targeting towards platelet-specific effects is in principle possible due to the specific substrate-kinase profile in platelets. To what extend these experimental drug candidates may have beneficial or adverse effects on platelets needs to be examined in experimental research. Furthermore, the integration of interaction information with drug data and GO terms allows the search for potential diagnostic markers.

There are also many limitations still present in platelet proteomics, which have to be overcome with the development of more optimized approaches. The platelet proteome is not yet complete and some false positive or false negative interactions may cause problems in the analysis of integrated networks (Goh, Lee et al. 2012).

6.2 Signal transduction in platelets

ADP modeling using integrated network analysis based on experimental phosphorylation and interaction data provides an integrated view on ADP signaling. The collaborative approach consisted of a Boolean model created by Mischnik et al and supported by experimental evidence from *PlateletWeb* interactions and phosphorylation events and the different phases could be represented as networks. Dynamical models of ADP signaling and cross-talk with other pathways has already been introduced by Wangorsch et al (Wangorsch, Butt et al. 2011). Similarly, SH2 domain binding proteins could be analyzed in a network dependent way. Missing links and interaction partners were thus revealed and the analyzed proteins could be connected with a known pathway to overview the whole system changes. Further investigation would involve excluding the tyrosine kinases and analyzing how the network signals without them.

This thesis also showed that visualizing subnetworks using the *PlateletWeb* knowledge base can be useful for creating novel hypotheses for the functional role of various aspects of the network, such as the case with the integrin signaling network. Integrin signaling has important clinical relevance and patients with functional defects in the integrin α2bβ3 receptor (Glanzmann thrombasthenia) may suffer severe bleeding disorders. Because of its crucial role in platelet activation, the α2bβ3 molecule is a well-known target for various pharmacological therapies in platelets and inhibitors such as abciximab are routinely involved in emergency coronary artery bypass grafting to reduce thrombosis risk (Seligsohn 2002). Based on the network analysis performed using integrated approaches and *PlateletWeb* data, we suggest new mechanisms for the regulation of the β3 cytoplasmic tail. The integrin receptor has already been introduced as a regulatory scaffold in previous studies (Oxley, Anthis et al. 2008; Shattil 2009). We propose a new putative regulatory mechanism based on the newly identified phosphorylation site Ser^{269} of the DOK1 molecule. DOK1 could act negatively on platelet function, as was already shown in Jurkat T-cells, where it inhibits PLCγ1 phosphorylation, Erk1/2 activation, and Ca^{2+} mobilization (Nemorin, Laporte et al. 2001). Furthermore, DOK proteins are involved in negative regulation of B-cell and T-cell signaling (Yamanashi, Tamura et al. 2000; Nemorin, Laporte et al. 2001). DOK1 is closely related to the DOK3 protein, which has also been indicated in integrin signaling. It has been proposed that DOK1 and DOK3 are negative regulators during α2bβ3 outside in signaling in the complex with SHIP-1 and Grb2 (Senis, Antrobus et al. 2009). Further investigation is needed to understand the functional role of DOK1 phosphorylation at Ser^{269} and its possible association with CLK1. Although the role of Ser^{269} phosphorylation on DOK1 and integrin signaling is not yet understood in detail, the predicted kinase for this site could be analyzed in experiments through inhibition by the chemical compound Debromohymenialdisine. There are reports about this kinase phosphorylating and activating the phosphatase PTPN1, which in turn facilitates platelet activity (Moeslein, Myers et al. 1999). Therefore, CLK1 may be involved in platelet activation, possibly through phosphorylating DOK1 on Ser^{269} and thus inhibiting the negative influence of DOK1 on the integrin receptor. Data extracted from network analysis is alone not enough to determine whether the DOK1 module might be a potential target for future antiplatelet therapy. Indeed, interest in this molecule has been raised and experiments on DOK1 regulation and kinetics have been already performed in platelets (Hughan and Watson 2007). In this study Watson et al demonstrated the expression of Dok1 in mouse and human platelets and showed that the protein is phosphorylated on a tyrosine residue after thrombin stimulation but not after GPVI or integrin α2bβ3 stimulation. They also demonstrated differential modes of regulation of Dok1 and Dok2 in platelets and introduced the idea of a role of Dok2 in outside-in signaling. Nonetheless, information about a serine phosphorylation in DOK1 and its possible functional significance has not yet been described. Dynamical modeling approaches lead to new hypotheses on the role of DOK1 during outside-in signaling, but there have not yet been any suggestions on its possible role in integrin inside-out signaling (Geier, Fengos et al. 2011).

Conclusive findings about the exact role of DOK1 in platelets have not yet emerged. Therefore, an in-depth experimental investigation of DOK1 signaling in platelets may be possible using the inhibitory compound of CLK1 and our resource provides required systems biological background for planning and evaluating such experiments. Through the analysis of integrin network with integrated network analysis, the advantages of using kinase predictions were emphasized in cases where experimental procedures are sparse or unavailable due to the lack of antibodies for phosphorylation analysis or an established mouse knockout model with platelet phenotype as in the case of DOK1(Hughan and Watson 2007). Thus, integrated network analysis coupled with kinase prediction can be a useful tool for functional exploration of new candidate proteins and phosphosites such as the phosphorylation of DOK1, which are not easily accessible experimentally.

6.3 Functional module detection based on functional similarity

By using the assembled human PPI network, I examined a module detection algorithm (Dittrich, Birschmann et al. 2008) enhanced with functional interaction information in two different cell systems. In a first approach, the performance of the algorithm was tested with and without functional interaction scores added to the interaction edges to underline the strength and advantages of using functional similarity. By analyzing a virus-infected gastric cell line dataset of 22 proteins, the algorithm extracted two different maximal scoring networks connecting the identified proteins. A closer look on the solution with interaction functional scores revealed that the module detection algorithm enhanced with functional edge scores provided connections between the proteins, which were confirmed by a higher number of experimental studies and therefore created a more reliable interpretation of the signaling changes in the whole protein network. Furthermore, the strengths of the algorithm were tested on a dataset from T-cells, where I used an additional option of extracting suboptimal solutions. The functional modules in these solutions represented typical functions for T-cells and confirmed that the algorithm enhanced with functional information can be used for characterization of various cell types and for decomposing a large dataset into its most relevant functional clusters.

We first tested the performance of our method on a small dataset derived from virus infected human gastric cells and compared it to a previous approach not integrating functional score information. There were clear differences in the topology and biological interpretation of the two resulting networks. While the main functional complex of keratin proteins was maintained in both, the connecting path to the other proteins in the sample was different. Although the phosphorylation of KRT18 at Ser^{53} by PRKCE is already known, the functional role of this phosphorylation has not yet been confirmed, it is speculated that it may play an in vivo role in filament reorganization (Omary, Baxter et al. 1992; Ku and Omary 1994), but it doesn’t explain the overall pathway and network context of the changed signaling in H9N2 infected gastric cells. When we focused on the keratin cluster in the solution including functional scores, it was associated with RAF1 and YWHAY (a 14-3-3 protein). Keratin 18 is known to regulate cell signaling via the association with 14-3-3 proteins and thereby with Raf1 kinase (Ku, Fu et al.

2004). During cell stress the binding between Raf1 and the keratin cluster is disrupted (Ku, Fu et al. 2004), which might be a useful hint in this case, as a virus infection causes stress to the gastric cells. KRT18 phosphorylation is reported to regulate various keratin functions including the binding to 14-3-3 proteins, involvement in the modulation of cell cycle progression and organizing keratin filaments (Ku, Azhar et al. 2002; Ku, Michie et al. 2002) along with a role in keratin protein turnover by ubiquitination (Ku and Omary 2000) or during apoptosis (Ku and Omary 2001). There is also evidence of a possible phosphorylation of KRT18 by Raf1, which causes the disruption in the complex (Ku, Fu et al. 2004). Therefore, our solution including functional scores reflects changes in signaling more accurately than the solution without functional scores. Further examples from the network confirm this finding. The interaction between pyruvate dehydrogenase alpha 1 (PDHA1) and pyruvate dehydrogenase kinase, isozyme 1 (PDK1) in the functional scores network has been identified in 24 studies and is a well-studied mechanism of cell signaling (Korotchkina and Patel 2001), while the interaction between PDHA1 and eukaryotic translation initiation factor 6 (EIF6) from the other network has just one broad interactome study as confirmation (Stelzl, Worm et al. 2005), where nothing specific is mentioned regarding this interaction. A well annotated interaction between guanine nucleotide binding protein (G protein), beta polypeptide 1 (GNB1) and guanine nucleotide binding protein (G protein), gamma 4 (GNG4) has been confirmed by three separate studies (Goddard, Ladds et al. 2006), the other alternative path (GNB1 to guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1 (GNB2L) (Dell, Connor et al. 2002)) is not as bmeaningful. For PHB, there is more information available for its interaction with RAF1 (Wang, Nath et al. 1999), while the interaction with SUMO4 is extracted only from a study with a list of proteins interacting with SMT3 suppressor of mif two 3 homolog 4 (SUMO4) (Guo, Han et al. 2005). The same holds true for the interaction between galectin-1 (LGALS1) and HRAS which has been identified as direct interaction and found to play an important role in mediating Ras membrane anchorage and cell transformation (Paz, Haklai et al. 2001). When we compared it with the interaction with survival of motor neuron protein interacting protein 1 (SIP1) (Park, Voss et al. 2001), SIP1 was part of the bigger SNM complex that associates with galectin and gemin4 for pre-mRNA splicing, but the study does not explicitly find a direct interaction between SIP1 and LGALS1.

These single interactions confirmed that the functional module detection algorithm complemented with functional interaction scores chooses interactions with more experimental evidence, which better explain the changes observed in virus infected cells.

In a second approach, the module detecting algorithm enhanced with functional information was tested on a T-cell dataset by using the option of extracting optimal and suboptimal solutions. Thus, the network could be decomposed into smaller characteristic functional modules. Focusing on particular functional modules helps to identify characteristic clusters for the cell type. In this case, specific functions to T-cells could be identified in the first suboptimal solution ("T-cell differentiation", "T-cell receptor signaling", "Signaling", "Kinase cascade", "Phosphorylation"). Using this option, large datasets can easily be split into smaller networks, which are better adapted for analysis. Enhancement of the module detection algorithm with semantic similarity data allows the detection of biologically significant and cell-specific modules from the sample. This method provides a smooth integration of data from various sources and is applicable to multiple proteome datasets for identification of functional modules.

The advantages of using network analysis in proteomics have already been thoroughly discussed in (Goh, Lee et al. 2012). According to the authors, alternative approaches are needed to complement existing methods to increase the comprehensiveness and precision of proteome coverage. Furthermore, network-based methods in proteomics can reduce the number of samples needed in a proteomic study. Currently, it is increasingly recognized that the understanding of properties that arise from whole-cell function require integrated, theoretical descriptions of the relationships between different cellular components (Albert 2005). An approach, which links exact solutions for maximum-scoring subnetworks with functional interaction data and PPI network background, makes analysis of qualitative proteomics much more comprehensive and focused on functionally relevant modules. Thus, characteristic features of the examined cell type can be analyzed and missing linking proteins in the network, which have not yet been identified due to limitations in the proteomics approaches can be added to the network solution. Although useful, this method also bears some limitations, as it is relying on the accuracy of the underlying PPI network, which may include false positive interactions (Goh, Lee et al. 2012). Other problems may also occur due to missing or insufficient functional annotations of the proteins in the Gene Ontology tree, which is then also included in the functional semantic similarity score.

False interpretation may therefore arise from false annotations, reflected in the final functional interaction score.

Future perspectives must also be considered, such as the network topology, which will change with the development of more exact and improved approaches for protein complex measurements. Quantitative proteomics tends to become more popular than qualitative approaches, because quantitative changes in the cell proteome are much more suitable for interpreting how the system reacts over time to different types of stimulation. As phosphoproteomics is also a growing field, site-specific quantitative phosphoproteomic studies need new methods for investigation of the kinase and phosphosite regulation in these datasets. Our method is particularly useful in such cases and there are already first methods developed to adapt it to phosphoprotemics (Chapter 6.4).

6.4 Response modules in quantitative phosphoproteomic data

Detection of signaling modules based on qualitative proteomics data was further extended to the investigation of quantitative phosphoproteomics data. In this thesis, I used an algorithm to detect signaling modules in large-scale phoshoproteomic networks. In contrast to previous methods, which mainly relied on gene expression data, our approach focuses on the analysis of cell-wide phosphorylation patterns. The integrated analysis combines protein-protein interaction (PPI) networks along with phosphoproteomic data to functionally describe signaling pathways and the change of information flow during various states of stimulation. To this end we used quantitative phosphoproteome data from embryonic stem cells after stimulation for differentiation for node scoring in networks derived from PPI data as well as kinase-substrate relationships. Subsequently, we searched for the maximum-scoring subnetwork using an exact algorithm to identify differentially phosphorylated signaling modules in cellular networks and thus obtained response modules which characterize precisely the regulation patterns observed during hESC differentiation.

The resulting networks enriched with both kinase-substrate information and phosphorylation sites illustrate the signaling flow, thereby reflecting changes during differentiation of hESCs and providing new insights on signaling mechanisms in differentially phosphorylated cellular networks. In summary, the algorithm reveals the integration of kinases derived from the human phosphoproteome significantly improves the systems biological understanding of signaling modules in a network context. Site specificity based on the maximum change in phosphorylation is crucial for identifying the most prominent effects of stimulation. When comparing these results to the results of a previous study on embryonic stem cells by van Hoof et al where they used kinase predictions to explain site regulation (Van Hoof, Dormeyer et al. 2010), our method stands out due to the experimental background and data used in our phosphoproteome assembly coupled with regulatory kinases extracted from the human phosphoproteome.

In a further analysis, I presented a tutorial for analyzing signaling modules from platelet quantitative phosphorylation data. Again, the algorithm proved useful for investigating the obtained proteins in the context of ADP signaling. Nonetheless, limitations are still present in platelet quantitative phosphoproteomics such as scarcity of supplied data and precision of the mass spectrometry methodology.

Future aspects of using functional module detection in quantitative phosphoproteomics data include the addition of gene ontology semantic similarity which can be included to the edges to enrich the functional information of the resulting modules as was already shown in Chapter 5.3 with qualitative proteomics data. Additionally, the algorithm can be enhanced by introducing directed edges to the input network for further analysis of the signal flow. Replicates can be considered for statistical validation of the investigated networks.

7 Conclusion and Outlook

This thesis presents first steps in the systems biological analysis of the platelet proteome and novel approaches for network analysis that improve the biological interpretation of identified proteins. Nonetheless, there are many challenges and future perspectives to be revealed in the field of platelet proteomics. The platelet proteome is far from complete, but the development of the *PlateletWeb* database considerably improves the wealth of information on platelets by integrating the most important available information and newly added phosphorylation sites. Quantitation is another important field which desperately needs attention and suffers from technical limitations. Determination of the protein levels and stoichiometry would significantly contribute to the understanding of dynamical changes in posttranslational platelet protein modifications. Based on this new data, hypothesis-generated research, such as the analysis of DOK1 signaling, can be further pursued in the future with quickly evolving mass-spectrometry techniques and lab experiments. One aspect of future development would be the addition of quantitative phosphorylation information to the platelet knowledge base, which would facilitate the investigation of platelet changes during various conditions. Dynamic information on the change of various proteins and kinases can be also included. Another aspect is the extension of the database to mouse proteins, which would help creating a platform for cross-species network analysis. The main idea behind this is that interactions existing in human cells but missing in mouse due to the lack of experimental data can be extrapolated based on the orthology of the interacting proteins. The same can be applied to human networks with missing components and edges, which are found in the mouse interactome. Thus, a multi-scale cross-species analysis will be available for platelet researchers, which would improve and simplify their efforts in understanding platelet signaling. As the mouse is a well-established model in cardio-vascular research, this repository would provide the platelet community with an invaluable tool for systems biological analysis of mouse as well as human platelets. With upcoming technologies and data optimization in mass-spectrometry, the database can be further optimized and supplied with platelet-specific information.

In the fields of network analysis, there is a lot more to be expected in future developments. The presented approach of combining functional information with PPI network context and proteomic data is just a first step towards reaching a better understanding of proteomics large-scale studies. Nonetheless, this shows once more that integrative analysis from different biological fields is essential to realize the whole potential of available information in an optimal way. And while proteomics data is often fractionized or insufficient for statistical validation, there are still ways to extract valuable information from the network context. However, it has to be noted, that our understanding of the interactome topology will change with the refinement of measurements and analytical methods. The human interactome is not merely a cloud with entities interacting with each other in a chaotic fashion. There are tendencies of creating a more appropriate modular view of the interactome with strongly interconnected protein complexes, which interact with other protein complexes using low affinity transient interactions combined with models in which also the three dimensional structure are considered (Stein, Mosca et al. 2011). There are already efforts involved in the analysis of protein complexes by integrating proteomics mass-spectrometry techniques with protein interaction networks, as reviewed by Stengel et al (Stengel, Aebersold et al. 2012), indicating that the future of network analysis also lies in understanding how interactions define the topology of the human interactome in a biological sense. Thus, network analysis will be more based on the biological and physical properties of proteins, rather than assuming only network characteristics.

Finally, collaborative efforts between technology, systems biology and computer science will be needed in future research to overcome the challenges of sparse data, technical difficulties and inconsecutive data analysis. Collaborations between different groups with special expertise will build the foundations of strong interdisciplinary research and ultimately provide results, which for now seem rather impossible to achieve. To accomplish this task optimization of data analysis and lab techniques will be of key importance. And although network analysis may face skepticism in the future, it is always good to remember that great ideas are not always recognized at first glance. One such example is the introduction of Gene Ontology as a hierarchically structured graph of gene functional annotations (Ashburner, Ball et al. 2000). Although it was considered with doubt in the beginning, now it has become an inseparable part of numerous bioinformatical and systems biological analyses. Following this example, the future of network analysis lies in the hands of people willing to think originally as the challenge is to comprehend the complex nature of cell function by analyzing all its components.

8 Supplementary materials

Supplemental Table 1. Number of proteins extracted from all studies

All platelet information sources are presented with the source of data, number of platelet proteins/transcripts and the fraction from which the proteins were extracted.

Study/Database	***Number of proteins***	***Fraction***
GPMBD 2011(Craig, Cortens et al. 2004)	2064	whole platelet
SAGE (Dittrich, Birschmann et al. 2006)	1964	whole platelet
Lewandrowski 2009 (Lewandrowski, Wortelkamp et al. 2009)	1269	membrane
Piersma 2009 (Piersma, Broxterman et al. 2009)	707	secretome
Haudek 2009 (Haudek, Slany et al. 2009)	671	whole platelet
Martens 2005 (Martens, Van Damme et al. 2005)	608	whole platelet
Garcia 2005 (Garcia, Smalley et al. 2005)	554	microparticles
Uniprot 2009 (2009)	538	undefined
Wong 2009 (Wong, McRedmond et al. 2009)	358	whole platelet
Garcia 2004 (García, Prabhakar et al. 2004)	305	whole platelet
Moebius 2005 (Moebius, Zahedi et al. 2005)	281	membrane
Thon 2008 (Thon, Schubert et al. 2008)	279	whole platelet
Zahedi 2008 (Zahedi, Lewandrowski et al. 2008)	277	phosphoproteome
HPRD (Keshava Prasad, Goel et al. 2009)	230	undefined
Maynard 2007 (Maynard, Heijnen et al. 2007)	206	alpha granules
Coppinger 2004 (Coppinger, Cagney et al. 2004)	182	secretome
Guerrier 2007 (Guerrier, Claverol et al. 2007)	176	whole platelet
Coppinger 2007 (Coppinger, Fitzgerald et al. 2007)	132	secretome
Gene RIF	120	undefined
Springer 2009 (Springer, Miller et al. 2009)	119	whole platelet
O'Neill 2002 (O'Neill, Brock et al. 2002)	116	whole platelet
Marcus 2000 (Marcus, Immler et al. 2000)	109	whole platelet
Garcia 2006 (Garcia, Senis et al. 2006)	77	whole platelet
Yu 2010 (Yu, Leng et al. 2010)	77	whole platelet
Thiele 2007 (Thiele, Steil et al. 2007)	29	whole platelet
Glenister 2008 (Glenister, Payne et al. 2008)	16	whole platelet

Supplemental Table 2. List of platelet human phosphatases.

The catalogue of human phosphatases was acquired from the Human Protein Phosphatases PCR Array (Quiagen) and complemented by manual search of phosphatases in our protein repository (see Materials and Methods). The total number of human phosphatases adds up to 191. Next, we analyzed and found that 73 phosphatases have evidence for expression in platelets.

Gene Symbol	*Gene ID*	*Description*
PPP3CA	5530	protein phosphatase 3, catalytic subunit, alpha isozyme
PPP3CB	5532	protein phosphatase 3, catalytic subunit, beta isozyme
PPP3CC	5533	protein phosphatase 3, catalytic subunit, gamma isozyme
PTPRC	5788	protein tyrosine phosphatase, receptor type, C
ACPP	55	acid phosphatase, prostate
PTPN11	5781	protein tyrosine phosphatase, non-receptor type 11
PTPN4	5775	protein tyrosine phosphatase, non-receptor type 4 (megakaryocyte)
PTPRB	5787	protein tyrosine phosphatase, receptor type, B
PTPN6	5777	protein tyrosine phosphatase, non-receptor type 6
PTPRA	5786	protein tyrosine phosphatase, receptor type, A
PTPN1	5770	protein tyrosine phosphatase, non-receptor type 1
PTPRG	5793	protein tyrosine phosphatase, receptor type, G
PTPN7	5778	protein tyrosine phosphatase, non-receptor type 7
PPP2CB	5516	protein phosphatase 2, catalytic subunit, beta isozyme
PPP6C	5537	protein phosphatase 6, catalytic subunit
MTM1	4534	myotubularin 1
TNS1	7145	tensin 1
PTPN12	5782	protein tyrosine phosphatase, non-receptor type 12
DUSP3	1845	dual specificity phosphatase 3
PTPRO	5800	protein tyrosine phosphatase, receptor type, O
PPP5C	5536	protein phosphatase 5, catalytic subunit
PPP2R4	5524	protein phosphatase 2A activator, regulatory subunit 4
PTPN9	5780	protein tyrosine phosphatase, non-receptor type 9
PTPRJ	5795	protein tyrosine phosphatase, receptor type, J
PPP3R1	5534	protein phosphatase 3, regulatory subunit B, alpha
PTP4A2	8073	protein tyrosine phosphatase type IVA, member 2
PTP4A1	7803	protein tyrosine phosphatase type IVA, member 1
PPP2R5A	5525	protein phosphatase 2, regulatory subunit B', alpha
PPP2R5C	5527	protein phosphatase 2, regulatory subunit B', gamma

Gene Symbol	***Gene ID***	***Description***
PTEN	5728	phosphatase and tensin homolog
PPP1R2	5504	protein phosphatase 1, regulatory (inhibitor) subunit 2
PPP1R12A	4659	protein phosphatase 1, regulatory (inhibitor) subunit 12A
PTPRK	5796	protein tyrosine phosphatase, receptor type, K
PPP1R7	5510	protein phosphatase 1, regulatory (inhibitor) subunit 7
DUSP2	1844	dual specificity phosphatase 2
DUSP5	1847	dual specificity phosphatase 5
PPP2R1B	5519	protein phosphatase 2, regulatory subunit A, beta
PPP1R9B	84687	protein phosphatase 1, regulatory (inhibitor) subunit 9B
PPP1R3D	5509	protein phosphatase 1, regulatory (inhibitor) subunit 3D
CDC14B	8555	CDC14 cell division cycle 14 homolog B (S. cerevisiae)
PPP1R12B	4660	protein phosphatase 1, regulatory (inhibitor) subunit 12B
PPM1B	5495	protein phosphatase, Mg2+/Mn2+ dependent, 1B
PPP2R2B	5521	protein phosphatase 2, regulatory subunit B, beta
MTMR12	54545	myotubularin related protein 12
PTPN18	26469	protein tyrosine phosphatase, non-receptor type 18 (brain-derived)
PTPRT	11122	protein tyrosine phosphatase, receptor type, T
DUSP23	54935	dual specificity phosphatase 23
MTMR14	64419	myotubularin related protein 14
ACP1	52	acid phosphatase 1, soluble
PPP1CC	5501	protein phosphatase 1, catalytic subunit, gamma isozyme
PPP2CA	5515	protein phosphatase 2, catalytic subunit, alpha isozyme
PPP2R5E	5529	protein phosphatase 2, regulatory subunit B', epsilon isoform
PPP2R2A	5520	protein phosphatase 2, regulatory subunit B, alpha
PPM1G	5496	protein phosphatase, Mg2+/Mn2+ dependent, 1G
SSH3	54961	slingshot homolog 3 (Drosophila)
SBF2	81846	SET binding factor 2
PPM1F	9647	protein phosphatase, Mg2+/Mn2+ dependent, 1F
PPP3R2	5535	protein phosphatase 3, regulatory subunit B, beta
PPP6R3	55291	protein phosphatase 6, regulatory subunit 3
PPP1R12C	54776	protein phosphatase 1, regulatory (inhibitor) subunit 12C
PPP1CB	5500	protein phosphatase 1, catalytic subunit, beta isozyme
DUSP26	78986	dual specificity phosphatase 26 (putative)
PPP1R16B	26051	protein phosphatase 1, regulatory (inhibitor) subunit 16B
PPP1R1C	151242	protein phosphatase 1, regulatory (inhibitor) subunit 1C

Gene Symbol	*Gene ID*	*Description*
PPP2R2D	55844	protein phosphatase 2, regulatory subunit B, delta
PPP1CA	5499	protein phosphatase 1, catalytic subunit, alpha isozyme
PPP2R1A	5518	protein phosphatase 2, regulatory subunit A, alpha
PDP1	54704	pyruvate dehyrogenase phosphatase catalytic subunit 1
PPM1A	5494	protein phosphatase, Mg2+/Mn2+ dependent, 1A
PPP1R14A	94274	protein phosphatase 1, regulatory (inhibitor) subunit 14A
ILKAP	80895	integrin-linked kinase-associated serine/threonine phosphatase
PPP6R1	22870	protein phosphatase 6, regulatory subunit 1
PPM1L	151742	protein phosphatase, Mg2+/Mn2+ dependent, 1L

Supplemental Table 3. Underrepresented pathways in platelets.

The Table represents pathways, which are significantly underrepresented in platelets, as they contain fewer platelet proteins than would be expected by chance. The underrepresentation was tested using the Fisher's Test for enrichment analysis (see Methods and Materials). Olfactory and neuronal receptors are lacking in the platelet cell, therefore there is a significant underrepresentation of the olfactory transduction and neuroactive ligand-receptor interaction pathways.

On the side of underrepresented pathways (Table S1) we find e.g. the "Neuroactive ligand-receptor interaction" (p-value = 2.06e-15) pathway which consists largely of neuronal receptors (220 of 318 proteins) and olfactory receptors; both classes of tissue-specific receptors, which are not expressed in platelets.

Pathway	*Total number of proteins*	*Number of Platelet proteins*	*Number of Non Platelet proteins*	*p-value*
Olfactory transduction	388	14	374	4.36E-50
Neuroactive ligand-receptor interaction	318	46	272	2.06E-15
Basal cell carcinoma	55	2	53	7.22E-07
Cytokine-cytokine receptor interaction	275	58	217	3.60E-06
Basal transcription factors	37	2	35	6.25E-04
Base excision repair	34	2	32	1.70E-03
Folate biosynthesis	64	9	55	2.37E-03
Aminoacyl-tRNA biosynthesis	71	11	60	2.81E-03
Glycosphingolipid biosynthesis - lacto and neolacto series	26	1	25	3.15E-03
DNA replication	36	3	33	3.15E-03
Intestinal immune network for IgA production	49	6	43	3.31E-03
Steroid hormone biosynthesis	56	8	48	4.20E-03
Hedgehog signaling pathway	56	8	48	4.20E-03
Autoimmune thyroid disease	54	8	46	7.10E-03

Supplemental Table 4. Proteins of the integrin signaling pathway.

The phosphorylation state of each protein is depicted in the right column.

Gene ID	*Gene Symbol*	*Description*
5908	RAP1B	RAP1B, member of RAS oncogene family
5906	RAP1A	RAP1A, member of RAS oncogene family
1796	DOK1	docking protein 1, 62kDa (downstream of tyrosine kinase 1)
5777	PTPN6	protein tyrosine phosphatase, non-receptor type 6
5336	PLCG2	phospholipase C, gamma 2 (phosphatidylinositol-specific)
5175	PECAM1	platelet/endothelial cell adhesion molecule
2533	FYB	FYN binding protein
7094	TLN1	talin 1
5781	PTPN11	protein tyrosine phosphatase, non-receptor type 11
3690	ITGB3	integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)
392	ARHGAP1	Rho GTPase activating protein 1
79930	DOK3	docking protein 3
10666	CD226	CD226 molecule
2317	FLNB	filamin B, beta
2316	FLNA	filamin A, alpha
9046	DOK2	docking protein 2, 56kDa
387	RHOA	ras homolog gene family, member A
6714	SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)
2534	FYN	FYN oncogene related to SRC, FGR, YES
5578, 5579, 5580, 5581, 5582, 5588, 5590	PRKCA, PRKCB, PRKCD, PRKCE, PRKCG, PRKCQ, PRKCZ	protein kinase C = PKC
1445	CSK	c-src tyrosine kinase
6093	ROCK1	Rho-associated, coiled-coil containing protein kinase 1
1195	CLK1	CDC-like kinase 1
6850	SYK	spleen tyrosine kinase
5566	PRKACA	protein kinase, cAMP-dependent, catalytic, alpha
5747	PTK2	PTK2 protein tyrosine kinase 2

Supplemental Table 5. Drugs of the integrin signaling pathway.

Associated drugs are listed along with drug type data and targeted proteins.

Associated drugs	***Drug type***	***Associated proteins***
13-Acetylphorbol	Experimental	PKC
2-Methyl-2,4-Pentanediol	Experimental	PRKACA
3,5-Diiodotyrosine	Experimental	PRKACA
3-pyridin-4-yl-1H-indazole	Experimental	PRKACA
5-benzyl-1,3-thiazol-2-amine	Experimental	PRKACA
Abciximab	Approved	ITGB3
Adenosine-5'-Diphosphate	Experimental	PTK2
Antithymocyte globulin	Approved	ITGB3
Balanol	Experimental	PRKACA
Citric Acid	Experimental	SRC
Cysteine Sulfenic Acid	Experimental	SRC
Dasatinib	Approved	FYN,SRC
Debromohymenialdisine	Experimental	CLK1
Dodecane-Trimethylamine	Experimental	PTPN11
Eptifibatide	Approved	ITGB3
Guanosine-5'-Diphosphate	Experimental	RHOA
Hydroxyfasudil	Experimental	PRKACA,ROCK1
Malonic acid	Experimental	SRC
N6-Benzyl Adenosine-5'-Diphosphate	Experimental	SRC
N-Octane	Experimental	PRKACA
O-Phosphoethanolamine	Experimental	PKC
Oxalic Acid	Experimental	SRC
Pentanal	Experimental	PRKACA
Phenylphosphate	Experimental	SRC
Phosphatidylserine	Approved	PKC
Phosphonoserine	Experimental	PRKACA,PKC
Phosphonothreonine	Experimental	PRKACA,PKC
Phosphonotyrosine	Experimental	SRC
Purvalanol A	Experimental	SRC
Staurosporine	Experimental	CSK,PKC,SYK
Sti-571	Experimental	SYK
Tirofiban	Approved	ITGB3
Vitamin E	Approved	PKC
Myristic acid	Experimental	PRKACA

Supplemental Table 6. All interactions available in the H9N2-virus infected cells solution without using functional interaction scores

Interactor 1	*Interactor 2*	*Detection Method*	*Pubmed IDs*	*Functional Interaction Score*
ACTB	PRKCD	in vitro	11415434	-2.5870518039347
ACTC1	CAPN1	yeast 2-hybrid	12358155	-2.84860074400133
ACTC1	HSPB1	in vivo	12087068	-1.89012302481934
ACTC1	PRKCE	in vivo	11968018	-2.65903506166956
APRT	IKBKG	NA	20098747	-1.58330872459433
C1QBP	PRKCD	in vivo;in vitro;yeast 2-hybrid	10831594	-2.4652260164825
CAPN1	ECHS1	yeast 2-hybrid	12358155	-2.83061924922273
CLIC1	SUMO4	in vivo	16236267	-2.10302004720975
GNB1	GNB2L1	in vitro	12359736	-2.68558191879177
EIF6	PDHA1	yeast 2-hybrid	16169070	-2.76371572785384
EIF6	GNB2L1	in vivo;in vitro;yeast 2-hybrid	14654845	-2.8006912417969
KRT1	PRKCE	in vivo	11897493	-2.75635028374924
KRT14	TCHP	yeast 2-hybrid	15731016	-1.91372994445305
KRT15	KRT18	yeast 2-hybrid	16189514	-1.28563526173082
KRT15	KRT19	yeast 2-hybrid	16189514	-1.10651104507118
KRT15	KRT81	yeast 2-hybrid	16189514	-1.12553141762592
KRT16	TCHP	yeast 2-hybrid	15731018	-1.88253735642867
KRT18	PPM1B	NA	17353931	-2.84000218388119
KRT18	PRKCE	in vitro	7523419,1374067, 15368451	-2.83321042541182
KRT18	TCHP	in vivo;in vitro;yeast 2-hybrid	15731014	-1.94116944361335
RPSA	SUMO4	in vivo	16236267	-2.10302004720975
LGALS1	SIP1	NA	11522829	-2.57470679834156
PDHA1	PDHB	in vivo	7864652	-0.135635033356706
PHB	SUMO4	in vivo	16236267	-2.10302004720975
PPIA	S100A8	yeast 2-hybrid	16169070	-2.37531768105933
PPM1B	S100A8	NA	17353931	-2.68820198109401
PPM1B	IKBKG	in vivo	14585847	-2.76573630495227
PPM1B	ISG15	NA	16884686	-2.87018692813316
PRKCD	GNB2L1	in vivo	11884618	-2.66714067777954

Interactor 1	***Interactor 2***	***Detection Method***	***Pubmed IDs***	***Functional Interaction Score***
PRKCE	GNB2L1	in vivo;in vitro	11956211,11709417	-2.62048881751121
RPS12	IKBKG	NA	20098747	-2.1208040568425
S100A8	NUAK1	NA	17353931	-2.52288288299586
SIP1	IKBKG	NA	20098747	-2.13278788728704
NUAK1	KRT77	NA	17353931	-2.10302004720975
GNB2L1	SUMO4	in vivo	16236267	-2.10302004720975
PRDX4	SUMO4	in vivo	16236267	-2.10302004720975

Supplemental Table 7. All interactions available in the H9N2-virus infected cells solution with functional interaction scores

Interactor 1	*Interactor 2*	*Detection Method*	*Pubmed IDs*	*Functional Interaction Score*
ACTB	ACTG1	yeast 2-hybrid	16189514	-0.515204327071414
ACTG1	SUMO4	in vivo	16236267	-2.10302004720975
AKT1	HSPB1	NA;in vivo;in vitro	11042204	-1.91239203821897
AKT1	PDK1	NA	15678105	-2.1270848204599
AKT1	RAF1	in vivo;in vitro	10576742,19058874, 10576742,11971957, 11997508,18669648, 18691976,18767875, 20230923,20068231	-1.64350182724194
AKT1	NUAK1	in vivo;in vitro	Not available	-1.76408895217485
APRT	IKBKG	NA	20098747	-1.58330872459433
C1QBP	YWHAB	in vivo	15324660	-2.82690595138411
CLIC1	SUMO4	in vivo	16236267	-2.10302004720975
ECHS1	EPB41	NA	17353931	-2.87873895850523
EPB41	YWHAB	yeast 2-hybrid	16368544	-2.81789002794067
GNB1	GNG4	in vivo;in vitro;yeast 2-hybrid	7665596,8636150, 16884933	-1.28748888738928
GNG4	RAF1	in vivo;in vitro	7782277	-2.36931223148317
HRAS	LGALS1	in vivo;in vitro	11709720	-2.32421982072048
HRAS	RAF1	in vivo;in vitro;yeast 2-hybrid	8530446,8911690, 8035810,9099670, 7730360,8332187, 16301319,15688026, 8332195,9154803, 9261098	-0.426130720732903
HSPA4	RAF1	in vivo;in vitro	16093354	-1.78816481515568
HSPA4	ISG15	NA	16884686	-2.80297571892739
JAK2	PPIA	in vivo	12668872	-2.60503512103199
JAK2	RAF1	in vivo;in vitro	8876196,10205168, 9689060,11134016, 10205168,8876196, 9689060,11134016,	-1.44417841580181

Interactor 1	***Interactor 2***	***Detection Method***	***Pubmed IDs***	***Functional Interaction Score***
			8876196	
JAK2	IKBKG	NA	20098747	-2.01133811929412
KRT1	KRT5	in vivo	11591653	-1.33826683895676
KRT5	KRT14	in vitro	8636216	-0.279920342893604
KRT5	KRT18	in vivo	9786957	-1.28563526173082
KRT5	TCHP	yeast 2-hybrid	15731015	-1.8215822131448
KRT15	KRT18	yeast 2-hybrid	16189514	-1.28563526173082
KRT15	KRT19	yeast 2-hybrid	16189514	-1.10651104507118
KRT15	KRT81	yeast 2-hybrid	16189514	-1.12553141762592
KRT16	TCHP	yeast 2-hybrid	15731018	-1.88253735642867
KRT18	YWHAB	in vivo	9524113	-2.26154453575568
RPSA	SUMO4	in vivo	16236267	-2.10302004720975
PDHA1	PDHB	in vivo	7864652	-0.135635033356706
PDHA1	PDK1	in vitro	12676647,11486000, 11485553	-0.045461107070217
PHB	RAF1	in vivo;in vitro	10523633	-2.39681571732858
PHB	SUMO4	in vivo	16236267	-2.10302004720975
RAF1	YWHAB	in vivo;in vitro	8702721,7644510	-0.983730579335193
RPS12	IKBKG	NA	20098747	-2.1208040568425
NUAK1	KRT77	NA	17353931	-2.10302004720975
PRDX4	SUMO4	in vivo	16236267	-2.10302004720975

Supplemental Table 8. Functional enrichment analysis of T-cells solutions without the use of functional interaction scores

A) BP of Optimal solution

GO-ID	*p-value*	*corr p-value*	*x*	*n*	*X*	*N*	*Description*
51130	3.5247E-6	3.6904E-3	9	25	47	809	positive regulation of cellular component organization
31346	1.0332E-5	5.4089E-3	5	7	47	809	positive regulation of cell projection organization
6928	1.6382E-5	5.7174E-3	12	53	47	809	cellular component movement
43434	2.6354E-5	6.8980E-3	5	8	47	809	response to peptide hormone stimulus
51128	6.3573E-5	1.2332E-2	12	60	47	809	regulation of cellular component organization
30036	7.5405E-5	**1.2332E-2**	10	43	47	809	**actin cytoskeleton organization**
30029	9.3375E-5	**1.2332E-2**	10	44	47	809	**actin filament-based process**
48856	9.4225E-5	1.2332E-2	19	140	47	809	anatomical structure development
60491	1.8443E-4	2.1456E-2	3	3	47	809	regulation of cell projection assembly
48731	2.1022E-4	2.2010E-2	17	123	47	809	system development
44087	3.0812E-4	2.7222E-2	7	25	47	809	regulation of cellular component biogenesis
31344	3.1200E-4	2.7222E-2	5	12	47	809	regulation of cell projection organization
7275	3.6928E-4	2.9175E-2	18	141	47	809	multicellular organismal development
9888	4.0810E-4	2.9175E-2	8	34	47	809	tissue development
7010	4.1799E-4	2.9175E-2	10	52	47	809	**cytoskeleton organization**
23034	4.8657E-4	3.0354E-2	13	84	47	809	intracellular signaling pathway
50793	4.9285E-4	3.0354E-2	10	53	47	809	regulation of developmental process
50896	6.9567E-4	3.9766E-2	23	217	47	809	response to stimulus
45596	7.2164E-4	3.9766E-2	5	14	47	809	negative regulation of cell differentiation

B) BP of first Suboptimal solution

GO-ID	***p-value***	***corr p-value***	***x***	***n***	***X***	***N***	***Description***
6446	2.9323E-5	1.1275E-2	5	8	48	809	regulation of translational initiation
10468	3.3177E-5	1.1275E-2	17	105	48	809	regulation of gene expression
44419	4.6148E-5	1.1275E-2	12	57	48	809	interspecies interaction between organisms

C) BP of second Suboptimal solution

GO-ID	***p-value***	***corr p-value***	***x***	***n***	***X***	***N***	***Description***
51640	3.5241E-5	1.9669E-2	6	13	47	809	organelle localization
6890	4.8190E-5	1.9669E-2	4	5	47	809	retrograde vesicle-mediated transport, Golgi to ER
48193	6.2178E-5	1.9669E-2	7	20	47	809	Golgi vesicle transport
51656	1.9027E-4	4.4762E-2	5	11	47	809	establishment of organelle localization
16050	3.0915E-4	4.4762E-2	4	7	47	809	vesicle organization
23033	3.2416E-4	4.4762E-2	16	115	47	809	**signaling pathway**
23052	3.3460E-4	4.4762E-2	20	166	47	809	**signaling**
6996	5.4909E-4	4.4762E-2	15	108	47	809	organelle organization
50789	5.5103E-4	4.4762E-2	34	392	47	809	regulation of biological process
280	5.9195E-4	4.4762E-2	4	8	47	809	nuclear division
7067	5.9195E-4	4.4762E-2	4	8	47	809	mitosis
8104	5.9436E-4	4.4762E-2	14	97	47	809	protein localization
48194	7.0752E-4	4.4762E-2	3	4	47	809	Golgi vesicle budding
48205	7.0752E-4	4.4762E-2	3	4	47	809	COPI coating of Golgi vesicle
48200	7.0752E-4	4.4762E-2	3	4	47	809	Golgi transport vesicle coating
65007	8.2297E-4	4.6794E-2	35	417	47	809	biological regulation
15031	8.8023E-4	4.6794E-2	13	89	47	809	protein transport
51641	9.1815E-4	4.6794E-2	14	101	47	809	cellular localization
45184	9.8511E-4	4.6794E-2	13	90	47	809	establishment of protein localization
87	1.0201E-3	4.6794E-2	4	9	47	809	M phase of mitotic cell cycle
279	1.0355E-3	4.6794E-2	5	15	47	809	M phase

Supplemental Table 9. Functional enrichment analysis of T-cells solutions using interaction scores

A) BP of Optimal solution

GO-ID	***p-value***	***corr p-value***	***x***	***n***	***X***	***N***	***Description***
44260	2.6174E-14	6.9100E-12	44	314	49	809	cellular macromolecule metabolic process
43170	9.7766E-13	1.2905E-10	44	341	49	809	macromolecule metabolic process
31145	3.7222E-8	1.6378E-6	12	31	49	809	**anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process**
51352	3.7222E-8	1.6378E-6	12	31	49	809	**negative regulation of ligase activity**
51444	3.7222E-8	1.6378E-6	12	31	49	809	**negative regulation of ubiquitin-protein ligase activity**
51436	3.7222E-8	1.6378E-6	12	31	49	809	**negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle**
51439	5.6960E-8	1.6708E-6	12	32	49	809	**regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle**
31397	5.6960E-8	1.6708E-6	12	32	49	809	**negative regulation of protein ubiquitination**
51437	5.6960E-8	1.6708E-6	12	32	49	809	positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle
44267	1.1686E-7	2.3859E-6	32	241	49	809	cellular protein metabolic process
51351	1.2652E-7	2.3859E-6	12	34	49	809	positive regulation of ligase activity
51340	1.2652E-7	2.3859E-6	12	34	49	809	regulation of ligase activity
51443	1.2652E-7	2.3859E-6	12	34	49	809	**positive regulation of ubiquitin-protein ligase activity**
51438	1.2652E-7	2.3859E-6	12	34	49	809	regulation of ubiquitin-protein ligase activity
31398	1.8412E-7	3.2406E-6	12	35	49	809	positive regulation of protein ubiquitination
31396	2.6411E-7	4.2358E-6	12	36	49	809	regulation of protein ubiquitination
44238	2.7276E-7	4.2358E-6	44	460	49	809	primary metabolic process
43161	3.7378E-7	5.0376E-6	12	37	49	809	proteasomal ubiquitin-dependent protein catabolic process
10498	3.7378E-7	5.0376E-6	12	37	49	809	proteasomal protein catabolic process
278	3.8164E-7	5.0376E-6	13	44	49	809	mitotic cell cycle
43489	6.6856E-7	8.0227E-6	5	5	49	809	RNA stabilization
48255	6.6856E-7	8.0227E-6	5	5	49	809	mRNA stabilization

GO-ID	***p-value***	***corr p-value***	***x***	***n***	***X***	***N***	***Description***
16071	9.2121E-7	1.0574E-5	11	33	49	809	**mRNA metabolic process**
6397	1.0193E-6	1.1166E-5	10	27	49	809	**mRNA processing**
44237	1.0574E-6	1.1166E-5	44	476	49	809	cellular metabolic process
8380	1.3013E-6	1.3213E-5	11	34	49	809	**RNA splicing**

B) BP of first suboptimal solution

GO-ID	***p-value***	***corr p-value***	***x***	***N***	***X***	***N***	***Description***
23033	1.7584E-8	1.0661E-5	23	115	50	809	**signaling pathway**
43687	2.1472E-8	1.0661E-5	19	79	50	809	**post-translational protein modification**
23014	5.2612E-8	1.0916E-5	9	16	50	809	**signal transmission via phosphorylation event**
7243	5.2612E-8	1.0916E-5	9	16	50	809	**intracellular protein kinase cascade**
7166	5.4965E-8	1.0916E-5	14	44	50	809	**cell surface receptor linked signaling pathway**
6464	1.8282E-7	3.0257E-5	19	89	50	809	protein modification process
43412	2.2240E-7	3.1549E-5	19	90	50	809	macromolecule modification
23052	3.0731E-7	3.8144E-5	26	166	50	809	**signaling**
30217	3.0231E-6	2.7250E-4	6	9	50	809	**T cell differentiation**
30098	3.0231E-6	2.7250E-4	6	9	50	809	**lymphocyte differentiation**
2429	4.2492E-6	2.7250E-4	5	6	50	809	**immune response-activating cell surface receptor signaling pathway**
2768	4.2492E-6	2.7250E-4	5	6	50	809	**immune response-regulating cell surface receptor signaling pathway**
50851	4.2492E-6	2.7250E-4	5	6	50	809	**antigen receptor-mediated signaling pathway**
50852	4.2492E-6	2.7250E-4	5	6	50	809	**T cell receptor signaling pathway**

C) BP of second suboptimal solution

GO-ID	***p-value***	***corr p-value***	***x***	***n***	***X***	***N***	***Description***
51128	3.0575E-14	2.5805E-11	22	60	49	809	regulation of cellular component organization
44087	8.7817E-11	3.7059E-8	13	25	49	809	regulation of cellular component biogenesis
6928	7.7897E-10	2.1915E-7	17	53	49	809	cellular component movement
30029	3.8848E-9	8.1969E-7	15	44	49	809	**actin filament-based process**
30036	2.9241E-8	4.3279E-6	14	43	49	809	**actin cytoskeleton organization**
32956	3.0767E-8	4.3279E-6	10	20	49	809	**regulation of actin cytoskeleton organization**
32970	5.6111E-8	6.7654E-6	10	21	49	809	regulation of actin filament-based process
22604	8.8259E-8	9.3113E-6	9	17	49	809	regulation of cell morphogenesis
31346	2.4436E-7	2.2916E-5	6	7	49	809	positive regulation of cell projection organization
51493	2.7195E-7	2.2952E-5	10	24	49	809	regulation of cytoskeleton organization
51130	4.3293E-7	3.0926E-5	10	25	49	809	positive regulation of cellular component organization
7010	4.4671E-7	3.0926E-5	14	52	49	809	cytoskeleton organization
30833	4.7635E-7	3.0926E-5	8	15	49	809	regulation of actin filament polymerization
43254	5.3068E-7	3.1993E-5	9	20	49	809	regulation of protein complex assembly
32271	9.0920E-7	4.6299E-5	8	16	49	809	regulation of protein polymerization
33043	9.2121E-7	4.6299E-5	11	33	49	809	regulation of organelle organization
50769	9.3257E-7	4.6299E-5	6	8	49	809	positive regulation of neurogenesis
31344	1.2271E-6	5.7539E-5	7	12	49	809	regulation of cell projection organization
8064	1.6389E-6	6.5792E-5	8	17	49	809	regulation of actin polymerization or depolymerization
30832	1.6389E-6	6.5792E-5	8	17	49	809	regulation of actin filament length
51960	1.6389E-6	6.5792E-5	8	17	49	809	regulation of nervous system development
65008	1.7150E-6	6.5792E-5	22	136	49	809	regulation of biological quality
22603	2.2464E-6	8.2435E-5	9	23	49	809	regulation of anatomical structure morphogenesis
10720	2.6693E-6	9.3870E-5	6	9	49	809	positive regulation of cell development
7159	3.8284E-6	1.2925E-4	5	6	49	809	**leukocyte cell-cell adhesion**

9 References

(2009). "The Universal Protein Resource (UniProt) 2009." Nucleic Acids Res **37**(Database issue): D169-174.

Akala, O. O. and M. F. Clarke (2006). "Hematopoietic stem cell self-renewal." Curr Opin Genet Dev **16**(5): 496-501.

Albert, R. (2005). "Scale-free networks in cell biology." J Cell Sci **118**(Pt 21): 4947-4957.

Albert, R., H. Jeong, et al. (2000). "Error and attack tolerance of complex networks." Nature **406**(6794): 378-382.

Almen, M. S., K. J. Nordstrom, et al. (2009). "Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin." BMC Biol **7**: 50.

Alonso, A., J. Sasin, et al. (2004). "Protein tyrosine phosphatases in the human genome." Cell **117**(6): 699-711.

Aloy, P. and R. B. Russell (2004). "Taking the mystery out of biological networks." EMBO Rep **5**(4): 349-350.

Altshuler, D., M. J. Daly, et al. (2008). "Genetic mapping in human disease." Science **322**(5903): 881-888.

Ando, S., K. Tanabe, et al. (1989). "Domain- and sequence-specific phosphorylation of vimentin induces disassembly of the filament structure." Biochemistry **28**(7): 2974-2979.

Anthis, N. J. and I. D. Campbell (2011). "The tail of integrin activation." Trends Biochem Sci **36**(4): 191-198.

Ashburner, M., C. A. Ball, et al. (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-29.

Avruch, J., A. Khokhlatchev, et al. (2001). "Ras activation of the Raf kinase: tyrosine kinase recruitment of the MAP kinase cascade." Recent Prog Horm Res **56**: 127-155.

Azuaje, F., H. Wang, et al. (2005). Ontology-driven similarity approaches to supporting gene functional assessment. Proceedings of The Eighth Annual Bio-Ontologies Meeting, Citeseer.

Barabasi, A. L. and R. Albert (1999). "Emergence of scaling in random networks." Science **286**(5439): 509-512.

Barabasi, A. L., N. Gulbahce, et al. (2011). "Network medicine: a network-based approach to human disease." Nat Rev Genet **12**(1): 56-68.

Barabasi, A. L. and Z. N. Oltvai (2004). "Network biology: understanding the cell's functional organization." Nat Rev Genet **5**(2): 101-113.

Barnouin, K. (2012). "Special issue in quantitative mass spectrometric proteomics." Amino Acids.

Bearer, E. L., J. M. Prakash, et al. (2002). "Actin dynamics in platelets." Int Rev Cytol **217**: 137-182.

Beisser, D., G. W. Klau, et al. (2010). "BioNet: an R-Package for the functional analysis of biological networks." Bioinformatics **26**(8): 1129-1130.

Benjamini, Y. and D. Yekutieli (2001). "The control of the false discovery rate in multiple testing under dependency." Annals of Statistics **29**: 1165--1188.

Bertalanffy, L. V. (1968). General Systems Theory: Foundations, Development, Applications, George Braziller.

Boisvert, F. M., Y. W. Lam, et al. (2010). "A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage." Mol Cell Proteomics **9**(3): 457-470.

Boja, E. S., D. Phillips, et al. (2009). "Quantitative mitochondrial phosphoproteomics using iTRAQ on an LTQ-Orbitrap with high energy collision dissociation." J Proteome Res **8**(10): 4665-4675.

Borroni, B., C. Agosti, et al. (2010). "Blood cell markers in Alzheimer Disease: Amyloid Precursor Protein form ratio in platelets." Exp Gerontol **45**(1): 53-56.

Boyanova, D., S. Nilla, et al. (2012). "PlateletWeb: a systems biologic analysis of signaling networks in human platelets." Blood **119**(3): e22-34.

Brill, L. M., W. Xiong, et al. (2009). "Phosphoproteomic analysis of human embryonic stem cells." Cell Stem Cell **5**(2): 204-213.

Broos, K., H. B. Feys, et al. (2011). "Platelets at work in primary hemostasis." Blood Rev **25**(4): 155-167.

Butland, G., J. M. Peregrin-Alvarez, et al. (2005). "Interaction network containing conserved and essential protein complexes in Escherichia coli." Nature **433**(7025): 531-537.

Calderwood, D. A., Y. Fujioka, et al. (2003). "Integrin beta cytoplasmic domain interactions with phosphotyrosine-binding domains: a structural prototype for diversity in integrin signaling." Proc Natl Acad Sci U S A **100**(5): 2272-2277.

Castagna, A., R. Polati, et al. (2012). "Monocyte/macrophage proteomics: recent findings and biomedical applications." Expert Rev Proteomics **9**(2): 201-215.

Cattaneo, M. and C. Gachet (2001). "The platelet ADP receptors." Haematologica **86**(4): 346-348.

Chien, C. T., P. L. Bartel, et al. (1991). "The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest." Proc Natl Acad Sci U S A **88**(21): 9578-9582.

Choudhary, C. and M. Mann (2010). "Decoding signalling networks by mass spectrometry-based proteomics." Nat Rev Mol Cell Biol **11**(6): 427-439.

Collins, M. O. and J. S. Choudhary (2008). "Mapping multiprotein complexes by affinity purification and mass spectrometry." Curr Opin Biotechnol **19**(4): 324-330.

Communi, D., V. Vanweyenberg, et al. (1997). "D-myo-inositol 1,4,5-trisphosphate 3-kinase A is activated by receptor activation through a calcium:calmodulin-dependent protein kinase II phosphorylation mechanism." EMBO J **16**(8): 1943-1952--.

Coppinger, J., D. J. Fitzgerald, et al. (2007). "Isolation of the platelet releasate." Methods Mol Biol **357**: 307--311.

Coppinger, J. A., G. Cagney, et al. (2004). "Characterization of the proteins released from activated platelets leads to localization of novel platelet proteins in human atherosclerotic lesions." Blood **103**(6): 2096--2104.

Craig, R., J. P. Cortens, et al. (2004). "Open source system for analyzing, validating, and storing protein identification data." J Proteome Res **3**(6): 1234-1242.

D'Alessandro, A., P. G. Righetti, et al. (2010). "The red blood cell proteome and interactome: an update." J Proteome Res **9**(1): 144-163.

Datta, A., F. Huber, et al. (2002). "Phosphorylation of beta3 integrin controls ligand binding strength." J Biol Chem **277**(6): 3943-3949.

Daub, H., J. V. Olsen, et al. (2008). "Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle." Mol Cell **31**(3): 438-448.

de Godoy, L. M. F., J. V. Olsen, et al. (2008). "Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast." Nature **455**(7217): 1251--1254.

del Pozo, A., F. Pazos, et al. (2008). "Defining functional distances over gene ontology." BMC Bioinformatics **9**: 50.

Dell, E. J., J. Connor, et al. (2002). "The betagamma subunit of heterotrimeric G proteins interacts with RACK1 and two other WD repeat proteins." J Biol Chem **277**(51): 49888-49895.

Dempfle, A., A. Scherag, et al. (2008). "Gene-environment interactions for complex traits: definitions, methodological requirements and challenges." Eur J Hum Genet **16**(10): 1164-1172.

Denis, M. M., N. D. Tolley, et al. (2005). "Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets." Cell **122**(3): 379-391.

Dittrich, M., I. Birschmann, et al. (2008). "Platelet protein interactions: map, signaling components, and phosphorylation groundstate." Arterioscler Thromb Vasc Biol **28**(7): 1326-1331.

Dittrich, M., I. Birschmann, et al. (2006). "Analysis of SAGE data in human platelets: features of the transcriptome in an anucleate cell." Thromb Haemost **95**(4): 643-651.

Dittrich, M. T., G. W. Klau, et al. (2008). "Identifying functional modules in protein-protein interaction networks: an integrated exact approach." Bioinformatics **24**(13): i223-231.

Dittrich, M. T., G. W. Klau, et al. (2008). "Identifying functional modules in protein-protein interaction networks: an integrated exact approach." Bioinformatics **24**(13): i223--i231.

Dorsam, R. T., S. Murugappan, et al. (2003). "Clopidogrel: Interactions with the P2Y12 Receptor and Clinical Relevance." Hematology **8**(6): 359-365.

Erler, J. T. and R. Linding "Network-based drugs and biomarkers." J Pathol **220**(2): 290-296.

Erler, J. T. and R. Linding (2010). "Network-based drugs and biomarkers." J Pathol **220**(2): 290-296.

Feijge, M. A., K. Ansink, et al. (2004). "Control of platelet activation by cyclic AMP turnover and cyclic nucleotide phosphodiesterase type-3." Biochem Pharmacol **67**(8): 1559-1567.

Fiehn, O., J. Kopka, et al. (2000). "Metabolite profiling for plant functional genomics." Nat Biotechnol **18**(11): 1157-1161.

Fields, S. and O. Song (1989). "A novel genetic system to detect protein-protein interactions." Nature **340**(6230): 245-246.

Frohlich, H., N. Speer, et al. (2007). "GOSim--an R-package for computation of information theoretic GO similarities between terms and gene products." BMC Bioinformatics **8**: 166.

Fujita, Y., H. Shirataki, et al. (1998). "Tomosyn: a syntaxin-1-binding protein that forms a novel complex in the neurotransmitter release process." Neuron **20**(5): 905-915.

Furie, B. and B. C. Furie (2008). "Mechanisms of thrombus formation." N Engl J Med **359**(9): 938-949.

Gandhi, T. K., J. Zhong, et al. (2006). "Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets." Nat Genet **38**(3): 285-293.

García, A., S. Prabhakar, et al. (2004). "Extensive analysis of the human platelet proteome by two-dimensional gel electrophoresis and mass spectrometry." Proteomics **4**(3): 656--668.

Garcia, A., Y. A. Senis, et al. (2006). "A global proteomics approach identifies novel phosphorylated signaling proteins in GPVI-activated platelets: involvement of G6f, a novel platelet Grb2-binding membrane adapter." Proteomics **6**(19): 5332-5343.

Garcia, B. A., D. M. Smalley, et al. (2005). "The platelet microparticle proteome." J Proteome Res **4**(5): 1516-1521.

Garcia, E. and D. Jay (2006). "[Platelet filamin: a cytoskeletal protein involved in cell signal integration and function]." Arch Cardiol Mex **76 Suppl 4**: S67-75.

Geier, F., G. Fengos, et al. (2011). "A computational analysis of the dynamic roles of talin, Dok1, and PIPKI for integrin activation." PLoS One **6**(11): e24808.

George, J. N., J. P. Caen, et al. (1990). "Glanzmann's thrombasthenia: the spectrum of clinical disease." Blood **75**(7): 1383-1395.

Giot, L., J. S. Bader, et al. (2003). "A protein interaction map of Drosophila melanogaster." Science **302**(5651): 1727-1736.

Glenister, K. M., K. A. Payne, et al. (2008). "Proteomic analysis of supernatant from pooled buffy-coat platelet concentrates throughout 7-day storage." Transfusion **48**(1): 99-107.

Goddard, A., G. Ladds, et al. (2006). "Identification of Gnr1p, a negative regulator of G alpha signalling in Schizosaccharomyces pombe, and its complementation by human G beta subunits." Fungal Genet Biol **43**(12): 840-851.

Goh, K. I., M. E. Cusick, et al. (2007). "The human disease network." Proc Natl Acad Sci U S A **104**(21): 8685-8690.

Goh, W. W., Y. H. Lee, et al. (2012). "How advancement in biological network analysis methods empowers proteomics." Proteomics **12**(4-5): 550-563.

Goldman, R. D., S. Khuon, et al. (1996). "The function of intermediate filaments in cell shape and cytoskeletal integrity." J Cell Biol **134**(4): 971-983.

Graham, G. J., Q. Ren, et al. (2009). "Endobrevin/VAMP-8-dependent dense granule release mediates thrombus formation in vivo." Blood **114**(5): 1083-1090.

Guerrier, L., S. Claverol, et al. (2007). "Exploring the platelet proteome via combinatorial, hexapeptide ligand libraries." J Proteome Res **6**(11): 4290--4303.

Guo, D., J. Han, et al. (2005). "Proteomic analysis of SUMO4 substrates in HEK293 cells under serum starvation-induced stress." Biochem Biophys Res Commun **337**(4): 1308-1318.

Guo, Z., L. Wang, et al. (2007). "Edge-based scoring and searching method for identifying condition-responsive protein-protein interaction sub-network." Bioinformatics **23**(16): 2121-2128.

Guzzi, P. H., M. Mina, et al. (2011). "Semantic similarity analysis of protein data: assessment with biological features and issues." Brief Bioinform.

Hart, G. T., A. K. Ramani, et al. (2006). "How complete are current yeast and human protein-interaction networks?" Genome Biol **7**(11): 120.

Hartwell, L. H., J. J. Hopfield, et al. (1999). "From molecular to modular cell biology." Nature **402**(6761 Suppl): C47-52.

Haudek, V. J., A. Slany, et al. (2009). "Proteome maps of the main human peripheral blood constituents." J Proteome Res **8**(8): 3834-3843.

Heinisch, G., W. Holzer, et al. (1996). "On the bioisosteric potential of diazines: diazine analogues of the combined thromboxane A2 receptor antagonist and synthetase inhibitor Ridogrel." J Med Chem **39**(20): 4058-4064.

Hennig, E. E., M. Mikula, et al. (2012). "Comparative kinome analysis to identify putative colon tumor biomarkers." J Mol Med (Berl) **90**(4): 447-456.

Hopkins, A. L. and C. R. Groom (2002). "The druggable genome." Nat Rev Drug Discov **1**(9): 727-730.

Hornbeck, P. V., I. Chabra, et al. (2004). "PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation." Proteomics **4**(6): 1551-1561.

Hughan, S. C. and S. P. Watson (2007). "Differential regulation of adapter proteins Dok2 and Dok1 in platelets, leading to an association of Dok2 with integrin alphaIIbbeta3." J Thromb Haemost **5**(2): 387-394.

Hunter, T. (2000). "Signaling--2000 and beyond." Cell **100**(1): 113-127.

Ideker, T., O. Ozier, et al. (2002). "Discovering regulatory and signalling circuits in molecular interaction networks." Bioinformatics **18 Suppl 1**: S233-240.

Ideker, T., V. Thorsson, et al. (2001). "Integrated genomic and proteomic analyses of a systematically perturbed metabolic network." Science **292**(5518): 929-934.

Ito, T., T. Chiba, et al. (2001). "A comprehensive two-hybrid analysis to explore the yeast protein interactome." Proc Natl Acad Sci U S A **98**(8): 4569--4574.

Jain, S. and G. D. Bader (2010). "An improved method for scoring protein-protein interactions using semantic similarity within the gene ontology." BMC Bioinformatics **11**: 562.

Jeong, H., S. P. Mason, et al. (2001). "Lethality and centrality in protein networks." Nature **411**(6833): 41-42.

Jin, J., T. M. Quinton, et al. (2002). "Adenosine diphosphate (ADP)-induced thromboxane A(2) generation in human platelets requires coordinated signaling through integrin alpha(IIb)beta(3) and ADP receptors." Blood **99**(1): 193-198.

Joo, S. J. (2012). "Mechanisms of Platelet Activation and Integrin alphaIIbeta3." Korean Circ J **42**(5): 295-301.

Jorgensen, C. and M. Locard-Paulet (2012). "Analysing signalling networks by mass spectrometry." Amino Acids.

Junker, B. H. and F. Schreiber (2008). Analysis of Biological Networks, Wiley-Interscience.

Kanehisa, M. (2002). "The KEGG database." Novartis Found Symp **247**: 91-101; discussion 101-103, 119-128, 244-152.

Katagiri, F. (2003). "Attacking complex problems with the power of systems biology." Plant Physiol **132**(2): 417-419.

Kerrigan, S. W. and D. Cox (2010). "Platelet-bacterial interactions." Cell Mol Life Sci **67**(4): 513-523.

Keshava Prasad, T. S., R. Goel, et al. (2009). "Human Protein Reference Database--2009 update." Nucleic Acids Res **37**(Database issue): D767-772.

Khalil, I., M. A. Brewer, et al. (2010). "The potential of biologic network models in understanding the etiopathogenesis of ovarian cancer." Gynecol Oncol **116**(2): 282-285.

Khor, S. (2010). "Concurrency and network disassortativity." Artif Life **16**(3): 225-232.

Kitano, H. (2001). Foundations of Systems Biology, MIT Press.

Knox, C., V. Law, et al. (2011). "DrugBank 3.0: a comprehensive resource for 'omics' research on drugs." Nucleic Acids Res **39**(Database issue): D1035-1041.

Konig, S., M. Nimtz, et al. (2012). "Kinome analysis of receptor-induced phosphorylation in human natural killer cells." PLoS One **7**(1): e29672.

Korotchkina, L. G. and M. S. Patel (2001). "Site specificity of four pyruvate dehydrogenase kinase isoenzymes toward the three phosphorylation sites of human pyruvate dehydrogenase." J Biol Chem **276**(40): 37223-37229.

Krishnan, S., M. Gaspari, et al. (2011). "OFFgel-based multidimensional LC-MS/MS approach to the cataloguing of the human platelet proteome for an interactomic profile." Electrophoresis **32**(6-7): 686-695.

Krogh, A., B. Larsson, et al. (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." J Mol Biol **305**(3): 567-580.

Ku, N. O., S. Azhar, et al. (2002). "Keratin 8 phosphorylation by p38 kinase regulates cellular keratin filament reorganization: modulation by a keratin 1-like disease causing mutation." J Biol Chem **277**(13): 10775-10782.

Ku, N. O., H. Fu, et al. (2004). "Raf-1 activation disrupts its binding to keratins during cell stress." J Cell Biol **166**(4): 479-485.

Ku, N. O., S. Michie, et al. (2002). "Keratin binding to 14-3-3 proteins modulates keratin filaments and hepatocyte mitotic progression." Proc Natl Acad Sci U S A **99**(7): 4373-4378.

Ku, N. O. and M. B. Omary (1994). "Identification of the major physiologic phosphorylation site of human keratin 18: potential kinases and a role in filament reorganization." J Cell Biol **127**(1): 161-171.

Ku, N. O. and M. B. Omary (2000). "Keratins turn over by ubiquitination in a phosphorylation-modulated fashion." J Cell Biol **149**(3): 547-552.

Ku, N. O. and M. B. Omary (2001). "Effect of mutation and phosphorylation of type I keratins on their caspase-mediated degradation." J Biol Chem **276**(29): 26792-26798.

Lambert, M. P. (2011). "What to do when you suspect an inherited platelet disorder." Hematology Am Soc Hematol Educ Program **2011**: 377-383.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lappe, M. and L. Holm (2004). "Unraveling protein interaction networks with near-optimal efficiency." Nat Biotechnol **22**(1): 98-103.

Leslie, M. (2010). "Cell biology. Beyond clotting: the powers of platelets." Science **328**(5978): 562-564.

Lewandrowski, U., S. Wortelkamp, et al. (2009). "Platelet membrane proteomics: a novel repository for functional research." Blood **114**(1): e10-19.

Li, B., F. Luo, et al. (2010). Effectively Integrating Information Content and Structural Relationship to Improve the GO-based Similarity Measure Between Proteins. BIOCOMP.

Li, S., C. M. Armstrong, et al. (2004). "A map of the interactome network of the metazoan C. elegans." Science **303**(5657): 540-543.

Linding, R., L. J. Jensen, et al. (2007). "Systematic discovery of in vivo phosphorylation networks." Cell **129**(7): 1415-1426.

Linding, R., L. J. Jensen, et al. (2007). "Systematic discovery of in vivo phosphorylation networks." Cell **129**(7): 1415--1426.

Linding, R., L. J. Jensen, et al. (2008). "NetworKIN: a resource for exploring cellular phosphorylation networks." Nucleic Acids Res **36**(Database issue): D695--D699.

Liu, J., M. E. Fitzgerald, et al. (2006). "Bruton tyrosine kinase is essential for botrocetin/VWF-induced signaling and GPIb-dependent thrombus formation in vivo." Blood **108**(8): 2596-2603.

Liu, J., T. I. Pestina, et al. (2004). "The roles of ADP and TXA in botrocetin/VWF-induced aggregation of washed platelets." J Thromb Haemost **2**(12): 2213-2222.

Liu, M., A. Liberzon, et al. (2007). "Network-based analysis of affected biological processes in type 2 diabetes models." PLoS Genet **3**(6): e96.

Liu, N., W. Song, et al. (2008). "Proteomics analysis of differential expression of cellular proteins in response to avian H9N2 virus infection in human cells." Proteomics **8**(9): 1851-1858.

Ljubi?, I., R. Weiskircher, et al. (2006). An algorithmic framework for the exact solution of the prize-collecting Steiner tree problem. Mathematical Progamming, Series B.

Lord, P. W., R. D. Stevens, et al. (2003). "Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation." Bioinformatics **19**(10): 1275-1283.

Ludwig, R. J., J. E. Schultz, et al. (2004). "Activated, not resting, platelets increase leukocyte rolling in murine skin utilizing a distinct set of adhesion molecules." J Invest Dermatol **122**(3): 830-836.

Macek, B., M. Mann, et al. (2008). "Global and Site-Specific Quantitative Phosphoproteomics: Principles and Applications." Annu Rev Pharmacol Toxicol.

Machida, K., C. M. Thompson, et al. (2007). "High-throughput phosphotyrosine profiling using SH2 domains." Mol Cell **26**(6): 899--915.

MacKay, V. L., X. Li, et al. (2004). "Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone." Mol Cell Proteomics **3**(5): 478-489.

Maere, S., K. Heymans, et al. (2005). "BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks." Bioinformatics **21**(16): 3448-3449.

Maglott, D., J. Ostell, et al. (2007). "Entrez Gene: gene-centered information at NCBI." Nucleic Acids Res **35**(Database issue): D26-31.

Maier, T., M. Guell, et al. (2009). "Correlation of mRNA and protein in complex biological samples." FEBS Lett **583**(24): 3966-3973.

Maki, K., H. Arai, et al. (2004). "Leukemia-related transcription factor TEL is negatively regulated through extracellular signal-regulated kinase-induced phosphorylation." Mol Cell Biol **24**(8): 3227-3237.

Malik, R., R. Lenobel, et al. (2009). "Quantitative analysis of the human spindle phosphoproteome at distinct mitotic stages." J Proteome Res **8**(10): 4553-4563.

Mann, M., S. E. Ong, et al. (2002). "Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome." Trends Biotechnol **20**(6): 261-268.

Manning, G., D. B. Whyte, et al. (2002). "The protein kinase complement of the human genome." Science **298**(5600): 1912-1934.

Marcus, K., D. Immler, et al. (2000). "Identification of platelet proteins separated by two-dimensional gel electrophoresis and analyzed by matrix assisted laser desorption/ionization-time of flight-mass spectrometry and detection of tyrosine-phosphorylated proteins." Electrophoresis **21**(13): 2622--2636.

Martens, L., P. Van Damme, et al. (2005). "The human platelet proteome mapped by peptide-centric proteomics: a functional protein profile." Proteomics **5**(12): 3193-3204.

Maynard, D. M., H. F. G. Heijnen, et al. (2007). "Proteomic analysis of platelet alpha-granules using mass spectrometry." J Thromb Haemost **5**(9): 1945--1955.

McGall, G. H. and F. C. Christians (2002). "High-density genechip oligonucleotide probe arrays." Adv Biochem Eng Biotechnol **77**: 21-42.

Merico, D., D. Gfeller, et al. (2009). "How to visually interpret biological data using networks." Nat Biotechnol **27**(10): 921-924.

Miller, M. L., L. J. Jensen, et al. (2008). "Linear motif atlas for phosphorylation-dependent signaling." Sci Signal **1**(35): ra2.

Moebius, J., R. P. Zahedi, et al. (2005). "The human platelet membrane proteome reveals several new potential membrane proteins." Mol Cell Proteomics **4**(11): 1754--1761.

Moeslein, F. M., M. P. Myers, et al. (1999). "The CLK family kinases, CLK1 and CLK2, phosphorylate and activate the tyrosine phosphatase, PTP-1B." J Biol Chem **274**(38): 26697-26704.

Munoz, J., T. Y. Low, et al. (2011). "The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells." Mol Syst Biol **7**: 550.

Nacu, S., R. Critchley-Thorne, et al. (2007). "Gene expression network analysis and applications to immunology." Bioinformatics **23**(7): 850-858.

Navlakha, S. and C. Kingsford (2010). "The power of protein interaction networks for associating genes with diseases." Bioinformatics **26**(8): 1057-1063.

Nemorin, J. G., P. Laporte, et al. (2001). "p62dok negatively regulates CD2 signaling in Jurkat cells." J Immunol **166**(7): 4408-4415.

Newman, D. K., S. Hoffman, et al. (2002). "Nitration of PECAM-1 ITIM tyrosines abrogates phosphorylation and SHP-2 binding." Biochem Biophys Res Commun **296**(5): 1171-1179.

Nousiainen, M., H. H. W. Silljé, et al. (2006). "Phosphoproteome analysis of the human mitotic spindle." Proc Natl Acad Sci U S A **103**(14): 5391--5396.

O'Neill, E. E., C. J. Brock, et al. (2002). "Towards complete analysis of the platelet proteome." Proteomics **2**(3): 288--305.

Obara, Y., K. Labudda, et al. (2004). "PKA phosphorylation of Src mediates Rap1 activation in NGF and cAMP signaling in PC12 cells." J Cell Sci **117**(Pt 25): 6085-6094.

Olsen, J. V., B. Blagoev, et al. (2006). "Global, in vivo, and site-specific phosphorylation dynamics in signaling networks." Cell **127**(3): 635--648.

Olsen, J. V., M. Vermeulen, et al. (2010). "Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis." Sci Signal **3**(104): ra3.

Omary, M. B., G. T. Baxter, et al. (1992). "PKC epsilon-related kinase associates with and phosphorylates cytokeratin 8 and 18." J Cell Biol **117**(3): 583-593.

Oxley, C. L., N. J. Anthis, et al. (2008). "An integrin phosphorylation switch: the effect of beta3 integrin tail phosphorylation on Dok1 and talin binding." J Biol Chem **283**(9): 5420-5426.

Ozlu, N., B. Akten, et al. (2010). "Phosphoproteomics." Wiley Interdiscip Rev Syst Biol Med **2**(3): 255-276.

Padovani, A., B. Borroni, et al. (2001). "Platelet amyloid precursor protein forms in AD: a peripheral diagnostic tool and a pharmacological target." Mech Ageing Dev **122**(16): 1997-2004.

Park, J. W., P. G. Voss, et al. (2001). "Association of galectin-1 and galectin-3 with Gemin4 in complexes containing the SMN protein." Nucleic Acids Res **29**(17): 3595-3602.

Pascal, L. E., L. D. True, et al. (2008). "Correlation of mRNA and protein levels: cell type-specific gene expression of cluster designation antigens in the prostate." BMC Genomics **9**: 246.

Patrono, C. and B. Rocca (2012). "Aspirin and Other COX-1 Inhibitors." Handb Exp Pharmacol(210): 137-164.

Paz, A., R. Haklai, et al. (2001). "Galectin-1 binds oncogenic H-Ras to mediate Ras membrane anchorage and cell transformation." Oncogene **20**(51): 7486-7493.

Pelkmans, L. and A. Helenius (2002). "Endocytosis via caveolae." Traffic **3**(5): 311-320.

Persico, M., A. Ceol, et al. (2005). "HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms." BMC Bioinformatics **6 Suppl 4**: S21.

Pesquita, C., D. Faria, et al. (2008). "Metrics for GO based protein semantic similarity: a systematic evaluation." BMC Bioinformatics **9 Suppl 5**: S4.

Pham, A. and J. Wang (2007). "Bernard-Soulier syndrome: an inherited platelet disorder." Arch Pathol Lab Med **131**(12): 1834-1836.

Pieroni, E., S. de la Fuente van Bentem, et al. (2008). "Protein networking: insights into global functional organization of proteomes." Proteomics **8**(4): 799-816.

Piersma, S. R., H. J. Broxterman, et al. (2009). "Proteomics of the TRAP-induced platelet releasate." J Proteomics **72**(1): 91-109.

Preisinger, C., A. von Kriegsheim, et al. (2008). "Proteomics and phosphoproteomics for the mapping of cellular signalling networks." Proteomics **8**(21): 4402--4415.

Purvis, J. E., M. S. Chatterjee, et al. (2008). "A molecular signaling model of platelet phosphoinositide and calcium regulation during homeostasis and P2Y1 activation." Blood **112**(10): 4069-4079.

Qiu, Y. Q., S. Zhang, et al. (2009). "Identifying differentially expressed pathways via a mixed integer linear programming model." IET Syst Biol **3**(6): 475-486.

R Development Core Team (2011). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

Ramirez, F., G. Lawyer, et al. (2012). "Novel search method for the discovery of functional relationships." Bioinformatics **28**(2): 269-276.

Resnik, P. (1995). Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In Proceedings of the 14th International Joint Conference on Artificial Intelligence.

Rigbolt, K. T., T. A. Prokhorova, et al. (2011). "System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation." Sci Signal **4**(164): rs3.

Rotilio, D., A. Della Corte, et al. (2012). "Proteomics: bases for protein complexity understanding." Thromb Res **129**(3): 257-262.

Rowley, J. W., A. Oler, et al. (2011). "Genome wide RNA-seq analysis of human and mouse platelet transcriptomes." Blood.

Roy, K. K. and E. A. Sausville (2001). "Early development of cyclin dependent kinase modulators." Curr Pharm Des **7**(16): 1669-1687.

Rual, J. F., K. Venkatesan, et al. (2005). "Towards a proteome-scale map of the human protein-protein interaction network." Nature **437**(7062): 1173-1178.

Ruffner, H., A. Bauer, et al. (2007). "Human protein-protein interaction networks and the value for drug discovery." Drug Discov Today **12**(17-18): 709-716.

Sanchez, C., C. Lachaize, et al. (1999). "Grasping at molecular interactions and genetic networks in Drosophila melanogaster using FlyNets, an Internet database." Nucleic Acids Res **27**(1): 89-94.

Sauer, U., M. Heinemann, et al. (2007). "Genetics. Getting closer to the whole picture." Science **316**(5824): 550-551.

Schlicker, A., F. S. Domingues, et al. (2006). "A new measure for functional similarity of gene products based on Gene Ontology." BMC Bioinformatics **7**: 302.

Schlicker, A., T. Lengauer, et al. (2010). "Improving disease gene prioritization using the semantic similarity of Gene Ontology terms." Bioinformatics **26**(18): i561-567.

Schwartz, A. S., J. Yu, et al. (2009). "Cost-effective strategies for completing the interactome." Nat Methods **6**(1): 55-61.

Schwertz, H., N. D. Tolley, et al. (2006). "Signal-dependent splicing of tissue factor pre-mRNA modulates the thrombogenicity of human platelets." J Exp Med **203**(11): 2433-2440.

Scott, J., T. Ideker, et al. (2006). "Efficient algorithms for detecting signaling pathways in protein interaction networks." J Comput Biol **13**(2): 133-144.

Scott, M. S., T. Perkins, et al. (2005). "Identifying regulatory subnetworks for a set of genes." Mol Cell Proteomics **4**(5): 683-692.

Seligsohn, U. (2002). "Glanzmann thrombasthenia: a model disease which paved the way to powerful therapeutic agents." Pathophysiol Haemost Thromb **32**(5-6): 216-217.

Senis, Y. A., R. Antrobus, et al. (2009). "Proteomic analysis of integrin alphaIIbbeta3 outside-in signaling reveals Src-kinase-independent phosphorylation of Dok-1 and Dok-3 leading to SHIP-1 interactions." J Thromb Haemost **7**(10): 1718-1726.

Senzel, L., D. V. Gnatenko, et al. (2009). "The platelet proteome." Curr Opin Hematol **16**(5): 329-333.

Sevilla, J. L., V. Segura, et al. (2005). "Correlation between gene expression and GO semantic similarity." IEEE/ACM Trans Comput Biol Bioinform **2**(4): 330-338.

Shannon, P., A. Markiel, et al. (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome Res **13**(11): 2498-2504.

Shattil, S. J. (2009). "The beta3 integrin cytoplasmic tail: protein scaffold and control freak." J Thromb Haemost **7 Suppl 1**: 210-213.

Song, Y. F., Y. Qu, et al. (2011). "Cellular localization of debromohymenialdisine and hymenialdisine in the marine sponge Axinella sp. using a newly developed cell purification protocol." Mar Biotechnol (NY) **13**(5): 868-882.

Songyang, Z., Y. Yamanashi, et al. (2001). "Domain-dependent function of the rasGAP-binding protein p62Dok in cell signaling." J Biol Chem **276**(4): 2459-2465.

Springer, D. L., J. H. Miller, et al. (2009). "Platelet proteome changes associated with diabetes and during platelet storage for transfusion." J Proteome Res **8**(5): 2261--2272.

Stegner, D. and B. Nieswandt (2011). "Platelet receptor signaling in thrombus formation." J Mol Med (Berl) **89**(2): 109-121.

Stein, A., R. Mosca, et al. (2011). "Three-dimensional modeling of protein interactions and complexes is going 'omics." Curr Opin Struct Biol **21**(2): 200-208.

Stelzl, U., U. Worm, et al. (2005). "A human protein-protein interaction network: a resource for annotating the proteome." Cell **122**(6): 957-968.

Stengel, F., R. Aebersold, et al. (2012). "Joining forces: integrating proteomics and cross-linking with the mass spectrometry of intact complexes." Mol Cell Proteomics **11**(3): R111 014027.

Stumpf, M. P., T. Thorne, et al. (2008). "Estimating the size of the human interactome." Proc Natl Acad Sci U S A **105**(19): 6959-6964.

Tadokoro, S., T. Nakazawa, et al. (2011). "A potential role for alpha-actinin in inside-out alphaIIbbeta3 signaling." Blood **117**(1): 250-258.

Tan, C. S., B. Bodenmiller, et al. (2009). "Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases." Sci Signal **2**(81): ra39.

Tao, Y., L. Sam, et al. (2007). "Information theory applied to the sparse gene ontology annotation network to predict novel gene function." Bioinformatics **23**(13): i529-538.

Thiele, T., L. Steil, et al. (2007). "Profiling of alterations in platelet proteins during storage of platelet concentrates." Transfusion **47**(7): 1221-1233.

Thomson, J. A., J. Itskovitz-Eldor, et al. (1998). "Embryonic stem cell lines derived from human blastocysts." Science **282**(5391): 1145-1147.

Thon, J. N., P. Schubert, et al. (2008). "Comprehensive proteomic analysis of protein changes during platelet storage requires complementary proteomic approaches." Transfusion **48**(3): 425-435.

Uetz, P., L. Giot, et al. (2000). "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae." Nature **403**(6770): 623-627.

Uetz, P. and M. J. Pankratz (2004). "Protein interaction maps on the fly." Nat Biotechnol **22**(1): 43-44.

Ulitsky, I. and R. Shamir (2007). "Identification of functional modules using network topology and high-throughput data." BMC Syst Biol **1**: 8.

Ulitsky, I. and R. Shamir (2009). "Identifying functional modules using expression profiles and confidence-scored protein interactions." Bioinformatics **25**(9): 1158-1164.

Van Hoof, D., W. Dormeyer, et al. (2010). "Identification of cell surface proteins for antibody-based selection of human embryonic stem cell-derived cardiomyocytes." J Proteome Res **9**(3): 1610-1618.

Varga-Szabo, D., I. Pleines, et al. (2008). "Cell adhesion mechanisms in platelets." Arterioscler Thromb Vasc Biol **28**(3): 403-412.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.

Vidal, M., M. E. Cusick, et al. (2011). "Interactome networks and human disease." Cell **144**(6): 986-998.

von Mering, C., L. J. Jensen, et al. (2005). "STRING: known and predicted protein-protein associations, integrated and transferred across organisms." Nucleic Acids Res **33**(Database issue): D433-437.

Walther, T. C. and M. Mann (2010). "Mass spectrometry-based proteomics in cell biology." J Cell Biol **190**(4): 491-500.

Wang, S., N. Nath, et al. (1999). "Rb and prohibitin target distinct regions of E2F1 for repression and respond to different upstream signals." Mol Cell Biol **19**(11): 7447-7460.

Wang, Y. (2008). "Condition specific subnetwork identification using an optimization model." Lecture notes in Operations Res. **9**: 333-340.

Wang, Z., M. Gerstein, et al. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nat Rev Genet **10**(1): 57-63.

Wangorsch, G., E. Butt, et al. (2011). "Time-resolved in silico modeling of fine-tuned cAMP signaling in platelets: feedback loops, titrated phosphorylations and pharmacological modulation." BMC Syst Biol **5**: 178.

Watanabe, N., M. Broome, et al. (1995). "Regulation of the human WEE1Hu CDK tyrosine 15-kinase during the cell cycle." EMBO J **14**(9): 1878-1891.

Wegener, K. L., A. W. Partridge, et al. (2007). "Structural basis of integrin activation by talin." Cell **128**(1): 171-182.

White, F. M. (2008). "Quantitative phosphoproteomic analysis of signaling network dynamics." Curr Opin Biotechnol **19**(4): 404--409.

Wollscheid, B., J. D. Watts, et al. (2004). "Proteomics/genomics and signaling in lymphocytes." Curr Opin Immunol **16**(3): 337-344.

Wong, J. W. H., J. P. McRedmond, et al. (2009). "Activity profiling of platelets by chemical proteomics." Proteomics **9**(1): 40--50.

Woollard, P. M., N. A. Mehta, et al. (2011). "The application of next-generation sequencing technologies to drug discovery and development." Drug Discov Today **16**(11-12): 512-519.

Wu, Z., X. Zhao, et al. (2009). "Identifying responsive functional modules from protein-protein interaction network." Mol Cells **27**(3): 271-277.

Yamanashi, Y., T. Tamura, et al. (2000). "Role of the rasGAP-associated docking protein p62(dok) in negative regulation of B cell receptor-mediated signaling." Genes Dev **14**(1): 11-16.

Yildirim, M. A., K. I. Goh, et al. (2007). "Drug-target network." Nat Biotechnol **25**(10): 1119-1126.

Yin, H., J. Liu, et al. (2008). "Src family tyrosine kinase Lyn mediates VWF/GPIb-IX-induced platelet activation via the cGMP signaling pathway." Blood **112**(4): 1139-1146.

Yu, Y., T. Leng, et al. (2010). "Global analysis of the rat and human platelet proteome - the molecular blueprint for illustrating multi-functional platelets and cross-species function evolution." Proteomics **10**(13): 2444-2457.

Zahedi, R. P., U. Lewandrowski, et al. (2008). "Phosphoproteome of resting human platelets." J Proteome Res **7**(2): 526-534.

Zahedi, R. P., U. Lewandrowski, et al. (2008). "Phosphoproteome of resting human platelets." J Proteome Res **7**(2): 526--534.

Zhang, J., N. Naslavsky, et al. (2012). "Rabs and EHDs: alternate modes for traffic control." Biosci Rep **32**(1): 17-23.

Zheng, S. and Z. Zhao (2011). "GenRev: Exploring functional relevance of genes in molecular networks." Genomics.

Zhong, J., S. A. Krawczyk, et al. (2010). "Temporal profiling of the secretome during adipogenesis in humans." J Proteome Res **9**(10): 5228-5238.

10 Additional Material

10.1 List of Figures

10.2 List of Tables

10.3 List of abbreviations

ADP	Adenosine Diphosphate
AGC	AGC kinase family
AP-MS	affinity-purification mass-spectrometry
APP	Amyloid beta Precursor Protein
Arp2/3 complex	Actin-Related Proteins complex
ATP	Adenosine Triphosphate
Atypical	kinase family of atypical protein kinases
BINGO	Biological Network Gene Ontology Tool
BP	Biological Process
BTK	Bruton tyrosine kinase
BUM	Beta Uniform Mixture Model
Ca2+	Calcium
CAMK	Calcium-calmodulin-dependent protein kinases family
cAMP	cyclic adenosine monophosphate
CC	Cellular Component
CDK2	Cyclin-dependent Kinase 2
CK1	Casein Kinase 1 family
CLK1	Dual specificity protein kinase CLK1
CMGC	CMGC kinase family
COX1/2	cyclooxygenase isoforms 1 and 2
DAG	directed acyclic graph
DOK1	Docking Protein 1
DOK2	Docking Protein 2
DOK3	Docking Protein 3
EIF6	eukaryotic translation initiation factor 6
EPH receptor	ephrin receptor
ERK1/2/Erk1/2	Extracellular-Signal-Regulated Kinases = MAP kinases
FDR	False Discovery Rate
GABA	gamma-aminobutyric acid
GNB1	guanine nucleotide binding protein (G protein), beta polypeptide 1
GNB2L1	guanine nucleotide binding protein (G protein), beta polypeptide 2-like
GO	Gene Ontology
GPIBA	Glycoprotein Ib (platelet), alpha polypeptide
GPIIb/IIIA	Glycoprotein IIb/IIIA
GPMDB	Global Proteome Machine Database
GTP	Guanosine Triphosphate
hESC	hESC human Embryonic Stem Cell
HMG-CoA	3-hydroxy-3-methylglutaryl-coenzyme A
HPRD	Human Protein Reference Database

InsP3R	inositol trisphosphate receptor
IP3	Inositol 1,4,5-trisphosphate
IP4	Inositol tetraphosphate
IPI	International Protein Index
IRS	Insulin Receptor Substrate
ITGB3	Integrin beta 3
ITPR1	Inositol 1,4,5-trisphosphate receptor type 1
ITRAQ	isobaric tags for relative and absolute quantitation
KEGG	KEGG Kyoto Encyclopedia of Genes and Genomes
KRT1	Keratin 1
KRT1	keratin 1
KRT18	Keratin 18
KRT18	keratin 18
LGALS1	galectin-1
MAP kinase	Mitogen-activated protein kinase
MEK	Mitogen-activated protein kinase kinase
MEKK	Mitogen-activated protein kinase kinase kinase
MF	Molecular Function
MICA	Most Informative Common Ancestor
mRNA	messenger RNA
NCM	Non-Controlled Medium
NCM	non-controlled medium
NO	nitric oxide
NP-hard	non-deterministic polynomial-time hard
PCST	prize-collecting Steiner tree problem
PDHA1	pyruvate dehydrogenase alpha 1
PDK1	pyruvate dehydrogenase kinase, isozyme 1
PECAM-1	platelet/endothelial cell adhesion molecule 1
Perl	Practical Extraction and Reporting Language
PH	Pleckstrin domain
PHB	prohibitin
PKA	Protein Kinase A
PKC	Protein Kinase C
PKG	Protein Kinase G
PLC	Phospholipase C
PPI	Protein-Protein Interaction
PTB	Phosphotyrosine Binding Domain
PTK	Protein tyrosine kinase
PTM	Posttranslational Modification
PTPN1	protein tyrosine phosphatase, non-receptor type 1
PTP-PEST	Protein Tyrosine Phosphatase PEST
RAS	Rat sarcoma
RGC	RGC kinase family

RIAM	Rap1-GTP-interacting adaptor molecule
S/Ser	Serine
SAGE	Serial Analysis of Gene Expression
SDS-Page	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH2	Src Homology 2 domain 2
SH3	Src Homology 3 domain 3
SILAC	Stable isotope labeling by amino acids in cell culture
SIP1	survival of motor neuron protein interacting protein 1
SNAP	Soluble NSF Attachment Protein
SNARE	Soluble NSF Attachment Protein Receptor
SRC	short for „sarcoma“
STE	STE kinase family
SUMO4	SMT3 suppressor of mif two 3 homolog 4
T/Thr	Threonine
TF	Tissue Factor
TK	Tyrosine Kinase family
TKL	Tyrosine Kinase-like Kinase family
TM	transmembrane
TXA2	Thromboxane A2
VASP	Vasodilator-stimulated Phosphoprotein
vWF	von Willebrand Factor
Y/Tyr	Tyrosine
Y2H	yeast-two hybride

11 List of publications

Boyanova D., Nilla S., Birschmann I., Dandekar T., Dittrich M. (2012). "*PlateletWeb*: a systems biologic analysis of signaling networks in human platelets." Blood **119** (3): e22-34.

Nilla S., **Boyanova D**., Dandekar T., Müller T., Dittrich M. (2012). "Identifying functional modules in protein-protein interaction networks using exact solutions and semantic similarity." *(in preparation)*

Mischnik M., **Boyanova D**., Dittrich M., Wangorsch G., Hubertus K., Geiger J., Timmer J.,Dandekar T. “A Boolean view separates platelet activatory and inhibitory signalling as verified by phosphorylation monitoring including threshold behaviour and integrin modulation.” *(in preparation)*

12 Conference contributions

GTH 2010, Nürnberg

D. Boyanova, S. Nilla, I. Birschmann, U. Walter, T. Dandekar, M. Dittrich
"*PlateletWeb*: An integrated Systems Biology platform for the analysis of Platelet Signaling"
1st Joint Meeting GTH & NVTH, Nürnberg (February 24th - 27th), 2010

SBMC 2010, Freiburg

D. Boyanova, S. Nilla, I. Birschmann, U. Walter, T. Dandekar, M. Dittrich
"Unravelling Cellular Networks: A systems biological perspective on platelet signaling"
3rd Conference on Systems Biology of Mammalian Cells, Freiburg (June 3rd - 5th), 2010

G. Wangorsch, M. Mischnik, K. Glausauer, S. Nilla, **D. Boyanova**, A. Sickmann, M. Dittrich, J. Timmer, J. Geiger, T. Dandekar
"PGI2 and ADP P2Y12 receptor signaling: downstream events and crosstalk"
3rd Conference on Systems Biology of Mammalian Cells, Freiburg (June 3rd - 5th), 2010

Boston 2010

G. Wangorsch, **D. Boyanova**, S. Nilla, M. Dittrich, T. Dandekar
"Integrating platelet proteome, phosphoproteome and drug information for a systems biological analysis of pharmacological targets"
International Conference on Systems Biology of Human Disease 2010 in Boston, MA, USA (June 16th -18th), 2010

GTH 2011, Wiesbaden

D. Boyanova, S. Nilla, I. Birschmann, U. Walter, T. Dandekar, M. Dittrich
"Integrated data analysis of functional protein networks in human platelets"
55th Annual Meeting of the GTH, Wiesbaden (February 16th-19th), 2011

G. Wangorsch, S. Nilla, **D. Boyanova**, M. Dittrich, T. Dandekar
"Probing modulatory interactions of the vWF/GP1b signaling cascade"
55th Annual Meeting of the GTH, Wiesbaden (February 16th-19th), 2011

M. Dittrich, **D. Boyanova**, S. Nilla, M. Balz, T. Dandekar, I. Birschmann
"An Integrated network of hemostatic drugs and drug targets in human platelets"
55th Annual Meeting of the GTH, Wiesbaden (February 16th-19th), 2011

ISMB 2011, Vienna

D. Boyanova, S. Nilla, D. Beisser, G. Klau, T. Dandekar, T. Müller, M. Dittrich
"Integrated analysis of cellular networks using phosphoproteome data to
identify active signaling modules"

S. Nilla, **D. Boyanova**, D. Beisser, G. Klau, T. Dandekar, T. Müller, M. Dittrich
"Identifying functional modules in protein-protein interaction networks
based on semantic similarity using an exact approach"
19th Annual International Conference on Intelligent Systems for Molecular Biology, Vienna (June 16th – 19th), 2011

European Conference

A. Zeeshan, M. Saman, C. Liang, A. Cecil, G. Wangorsch, S. Nilla, **D. Boyanova**, M. Naseem, A. Fieselmann, M. Dittrich., T. Dandekar
"Intelligent Information Management for efficient computational biology"
ICT 2011 - Information and Networking Day - Intelligent Information Management Jean Monnet Conference Centre, Luxembourg (26 September 2011)

<interact> Symposium 2012, Munich

D. Boyanova, S. Nilla, I. Birschmann, T. Dandekar, M.Dittrich
"Systems biological analysis of the human platelet proteome network" (talk)
<interact> symposium 2012, Munich (March 29th – 30th), 2011

Platelet Symposium 2012, Boston, USA

M. Dittrich, C. Kuhlmann, D. Boyanova, T. Dandekar, C. Knabbe, I. Birschmann
"*PlateletWeb*: A systems biology workbench to investigate signaling networks in patients with platelet disorders"
Platelet symposium 2012, Boston (June 7th – 22th), 2012

SFB688/CHFC Joint Symposium 2012, Würzburg

G. Wangorsch, **D.Boyanova**, M.Dittrich, T. Dandekar
"Dynamics and signaling in platelets"
SFB688/CHFC Joint Symposium 2012, Würzburg (June 21th – 22th)

SBMC 2012, Leipzig

D. Boyanova, G. Wangorsch, M. Dittrich, T. Dandekar
„Platelet activation and inhibition: Revealing feedback loops and drug interaction networks“
Systems Biology of Mammalian Cells, Leipzig (July 9th – 11th)

13 Curriculum vitae

Curriculum vitae

Desislava Boyanova
Peter-Schneider-Str.1, 97074 Würzburg, Germany
Mail: boyanova.desislava@googlemail.com
Mobile: +49 176 32039489

Personal information

Name	Desislava Veselinova Boyanova
Date of Birth	06.04.1985
Place of birth	Plovdiv, Bulgaria
Nationality	Bulgarian

Education

Since 06/2009	University of Würzburg, Germany Department of Bioinformatics Intended degree: PhD (Dr. rer. nat.)
10/2007 – 06/2009	M.Sc. program in Biomedicine University of Wuerzburg; Graduated with distinction 1,3 (excellent)
10/2004 – 07/2007	B.Sc. program in Biomedicine University of Wuerzburg; Graduated with 2,1 (good)
09/1999 - 05/2004	Foreign Language School “Plovdiv” with extensive German classes and English as second language; DSD (Deutsches Sprachdiplom) Stufe II
09/1992 - 06/1999	Primary school „Aleko Konstantinov“ with special language classes in English and Russian

Research experience

Since 06/2009	PhD student at the Department of Bioinformatics, University of Wuerzburg, PhD thesis: „Analysis of platelets and other biological networks”
09/2008 - 06/2009	Master Student at the Department of Bioinformatics, University of Wuerzburg, Master thesis: “The virtual platelet: examining building blocks of its interactome”
02/2008 – 05/2008	Internship in the Department of Human Genetics, mRNA Analysis Topic: „Analysis of MCPH1-RNA expression“
03/2007 – 06/2007	Bachelor Student at the Rudolf-Virchow-Center Wuerzburg Bachelor Thesis: “Regulation of endothelial antigen expression by tumor cells”

Prizes and scholarships

2004-2009	Deutscher Akademischer Austauschdienst (DAAD) full scholarship for the whole period of education as Master of Science in Biomedicine

Teaching experience

Since 2009	Supervision of Diploma and Bachelor Students; Practical courses in Network Analysis and Microarray Analysis for Bachelor students (Biology and Biomedicine)

Technical and lab skills

Software	R, Perl, MySQL, HTML, PHP
Programs	Cytoscape, Microsoft Office, Adobe Illustrator, Adobe Photoshop, LaTeX
Lab skills	Network Analysis, Microarray Analysis, ELISA, Proteomics data analysis, Cytokine Arrays, Cell culture, Real-Time PCR, Northern Blot, Western Blot, Immunostaining, Flow cytometry

Language skills

Bulgarian	mother tongue
German	fluent, DSD (Deutsches Sprachdiplom)
English	fluent
Russian	basics
Norwegian	basics
Swedish	level A2
French	level A2

Publications

Boyanova Desislava, Nilla Santosh, Birschmann Ingvild, Dandekar Thomas, Dittrich Marcus (2012). "PlateletWeb: a systems biologic analysis of signaling networks in human platelets." Blood 119(3): e22-34.

Place, Date Signature