Modeling Cancer Metastasis using Global, Quantitative and Integrative Network Biology

Schoof, Erwin; Brunak, Søren; Linding, Rune; Erler, Janine

Publication date:
2014

Document Version
Peer reviewed version

Modeling Cancer Metastasis using Global, Quantitative and Integrative Network Biology

Erwin M. Schoof

January 30, 2014

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Contents

Preface

This thesis is submitted as a requirement for obtaining the Ph.D degree at the Center for Biological Sequence Analysis (CBS), Department of Systems Biology, at the Technical University of Denmark (DTU) and was funded by a DTU scholarship.

All the work was carried out at the Center for Biological Sequence Analysis under the supervision of Professor Rune Linding and Dr Janine T. Erler.

Lyngby, January 2014
Erwin M. Schoof

Abstract

In order to respond to alterations in its environment, a cell has to integrate multiple input-cues and modulate its signaling networks accordingly, in order to elicit a specific response such as proliferation or apoptosis. This process becomes significantly altered during cancer development, with genomic modifications giving rise to differential protein dynamics, ultimately resulting in disease. The exact molecular signaling networks underlying specific disease phenotypes remain elusive, as the definition thereof requires extensive analysis of not only the genomic and proteomic landscapes within a particular tumor, but also the phenotypic response to perturbations. Thus, there is a critical need for an integrative global approach, which assesses a biological system such as cancer from several molecular aspects in an un-biased fashion. This thesis summarizes the efforts that were undertaken as part of my PhD in an attempt to positively contribute to this fundamental challenge.

The thesis is divided into four parts. In Chapter I, we introduce the complexity of cancer, and describe some underlying causes and ways to study the disease from different molecular perspectives. There is a nearly infinite number of biological aspects that would need to be understood to enable comprehensive treatment regimens specific to each patient (i.e. personalized medicine). However, in the approaches outlined in this thesis, we chose metastasis as a key process for interrogating the clinical potential of targeting cancer networks using Network Biology. Technologies key to this, such as Mass Spectrometry (MS), Next-Generation Sequencing (NGS) and High-Content Screening (HCS) are briefly described. In Chapter II, we cover how signaling networks and mutational data can be modeled in order to gain a better understanding of molecular processes which are fundamental to tumorigenesis. In Article 1, we propose a novel framework for how cancer mutations can be studied by taking into account their effect at the protein network level. In Article 2, we demonstrate how global, quantitative data on phosphorylation dynamics can be generated using MS, and how this can be modeled using a computational framework for deciphering kinase-substrate dynamics. This framework is described in depth in Article 3, and covers the design of KinomeXplorer, which allows the prediction of kinases responsible for modulating observed phosphorylation dynamics in a given biological sample. In Chapter III, we move into Integrative Network Biology, where, by combining two fundamental technologies (MS & NGS), we can obtain more in-depth insights into the links between cellular phenotype and genotype. Article 4 describes the proof-of-principle concept of how one can look at DNA mutations and protein dynamics in an integrative fashion. This has, for example, allowed us to investigate how mutations at the DNA level are propagated at the proteome level. Article 5 demonstrates how by taking a global, multi-platform approach, combined with extensive computational analysis, it is possible to gain a better understanding of colorectal cancer metastasis, and obtain potential clinical benefits.
Chapter IV briefly summarizes the findings of the thesis and closes by proposing some future directions based on the work that was presented.

Overall, the thesis aims to demonstrate the value of deploying several experimental platforms, each studying a different biological aspect, combined with in-depth computational analysis, in order to shed light on the fundamental molecular processes which underlie a complex disease like cancer and provide possible avenues for therapeutic intervention.

Dansk Resumé

For at kunne respondere til ændringer i miljøet, en celle skal integrere adskillige input-signaler og herefter modulere dens signalnetværk for at fremkalde et specifikt respons som for eksempel celle-deling eller apoptose. Denne proces ændres signifikant under cancer udvikling, hvorved genetiske modifikationer giver anledning til forskelige protein dynamikker, hvilket ultimativt kan resultere i sygdom. De eksakte molekylære signalnetværker som ligger til grunde for specifikke sygdoms fænotyper er stadig uafklaret, idet definitionen deraf kræver ekstensiv analyse, ikke kun på gen- og protein niveau i en tumor, men også det fænotypiske respons til perturbationer. Der eksisterer derfor et kritisk behov for en integrativ global tilgang, som uvildigt vurderer et biologisk system fra flere molekylære aspekter. Denne afhandling opsummerer den indsats, der blev iværksat som en del af min ph.d., i et forsøg på at bidrage positivt til denne grundlæggende udfordring.

Afhandlingen er opdelt i 4 dele. I kapitel I, introducerer vi til kompleksiteten af cancer og beskriver nogle af de underlæggende årsager og metoder til at studere sygdomen fra forskellige molekylære perspektiver. På trods af det næsten uendelige antal af biologiske aspekter som skal beskrives for at opnå en succesfuld behandling, har vi valgt metastase som en nøgle proces til at studere dens kliniske potentiale ved at bruge "Network Biology". Teknologier som kan opfylde dette, for eksempel Masse Spektrometri (MS), Next-Generation Sequencing (NGS) og High-Content Screening (HCS), er beskrevet. I kapitel II, dækker vi hvordan signalnetværker og mutations data kan moduleres til at opnå en bedre forståelse for de molekylære processer, som er fundamentale i tumorudvikling. I artikel 1, beskrive vi en ny ramme for hvordan cancer mutationer kan studeres ved at tage deres effekt på protein netværks niveau i betragtning. I artikel 2, demonstrerer vi hvordan globale, kvantitative data på fosforylerings dynamikker kan genereres ved at buge MS, og hvordan dette kan blive moduleret ved at bruge informatik rammer til at bestemme kinasesubstrat dynamikker. Denne ramme er beskrevet dybdegående i artikel 3, og dækker design af KinomeXplorer, som muliggøre forudsigelsen af kinaser, der modulerer observeret fosforylerings dynamikker i en given biologisk prøve.
I kapitel III, bevæger vi os ind i integrativ Network Biology, hvor, ved at kombinerer to grundlæggende teknologier (MS & NGS), kan vi opnå mere dybdegående indsigt i etableringen af fænotype fra genotype. Artikel 4 beskriver beviset-af-princippet på hvordan man kan se på DNA mutationer og protein dynamikker i en integrativ facon. Dette har for eksempel gjort det muligt at undersøge hvordan mutationer på DNA niveauet føres videre på protein niveau. Artikel 5 demonstrer hvordan, ved at tage en global, multi-platform tilgang kombineret med ekstensiv informatik analyse, det er muligt at opnå en bedre forståelse af kolon-endetarms cancer metastase, og opnå potentielle kliniske fordele. Kapitel IV opsummerer resultaterne af denne afhandling og afsluttes ved at foreslå fremtidige retninger baseret på det præsenterede arbejde.

Overordnet, denne afhandling har til formål at demonstrere værdien af at bruge flere eksperimentelle platforme, hvoraf hver studerer et forskelligt biologisk aspekt, kombineret med dybdegående informatikanalyse til at belyse de grundlæggende molekylærer processer, som ligger til grunde for en kompleks sygdom som cancer og levere muligheder for terapeutisk intervention.

Acknowledgements:

The work that has gone into this thesis and PhD is the result of a great collective effort, involving an extensive amount of brilliant people, whom have both made the work possible and enjoyable. I would like to extend my eternal gratitude to the following people:

First of all, my supervisor Rune Linding. Throughout my PhD, you have given me the right mix of scientific and technical freedom and supervision. Together with my other supervisor, Janine Erler, you have both allowed me to pursue my own interests, and move from a primarily computational PhD I initially came to London for, to a more elaborate one involving lots of lab work, extremely exciting cancer biology and many fruitful collaborations.
Next, I can't thank my PhD bro Pau enough for all the fun, excitement, encouragement and support throughout the years. I can't imagine having gone through this without you, and the countless days working late (even till 3am doing tissue culture) listening to Flaix Ibiza I will never forget. I will also always be thankful for the number of times I was stuck in analysis and we were able to discuss the problems and come up with some concrete things to try and solve them- really incredible! I am sure we will stay in touch, and once in a while instead of having a conference call with Mario, have one with each other. Similarly, I want to thank Cristina for all the help, support and fun we had, and showing me how amazing it can be to teach others experimental protocols, data analysis steps etc., and for being such a quick learner. I am sure you are going to do great things with the data you're generating and can't wait to see what comes out!
I'd also like to say thanks to Tom for all his help throughout the years in the lab, as you have made learning new lab skills a fun process and have helped me come to terms with the foes and woes of doing practical lab work. I really feel I learnt a lot from you, ranging from general tissue culture skills, through coating tons of 15cm dishes with collagen, lysing bucket-loads of cells and getting on top of basic wet lab procedures such as Western Blots and mycoplasma checks. In line with this, I will also be forever grateful to Annie for being such a good teacher and showing me how to work sterile with mammalian cells and for being very patient with me in the beginning, when my practical lab experience was extremely limited. Also the other members of the ErlerLab deserve a special mention, as you made going into the lab a very fun and fruitful experience!
I also want to thank the members of the LindingLab, and especially James for all the help with the RNAi and inhibitor screening, Jesper for the help with all the modeling (fascinating what the world of physics / mathematics can do for biology!), Jinho for helping with the sequencing data and KinomeXplorer, Craig for loads of tips that make life in the lab much easier and experimental workflows more efficient, and Lene with taking over my lab manager duties and allowing me to fully focus on my projects. Thanks also go out to Franziska for being a great student and for helping out in the MS lab; it's been a pleasure teaching you the protocols and I'm sure your experience and independence will help you along the way during your PhD.
Next, I will always be grateful for the help Adrian Pasculescu has given on many of our projects, and your look on data analysis as 'simple mathematics' will always help me keep a smile on my face while being faced with a difficult computational challenge. I couldn't have done at least half the analyses without your input, and I really hope we can keep collaborating in the future.
Martin Lee Miller also deserves a special thanks due to all the help he has given with me getting on top of the NetPhorest code. I had a great time meeting you in New York and it's been very fruitful to have you with us during the development of the next-generation versions of NetPhorest and NetworKIN. I've also always felt very grateful towards the CBS system administration (John, Peter and Kristoffer especially), as you were always available in dire times of need and have taught me a lot about system administration and large computational infrastructures. You guys really understand the value of providing great support and I hope you will always keep up this great spirit! Likewise, I would also like to thank the CBS administration (especially Dorthe) for all the administrative help which I could not have been without throughout the years, and of course to Søren Brunak for taking us in from London and providing me with much needed support to keep my PhD going. Similarly, my

sincere thanks goes out to Susanne Brix Pedersen, for allowing me to use her lab facilities when our own had not yet been established, as this allowed me to keep the lab work going without much delay.
I would also like to thank ThermoFisher for their great support over the years, as setting up and running a mass spec lab would have been near impossible without this! Lars, Peter and Vlad have really helped me gain a better understanding of how a mass spectrometer and LC system work and how to practically operate and maintain them; for this I will forever be grateful!
A special thanks goes out to my family. My mom, Anja, for always being so positive and supportive, and being there in times of need, and my dad, Hans, for always being there for me too, and pushing me to pursue my ambitions and encouraging me to explore my scientific curiosity. I will also always be grateful to my sisters Helmi and Vera, as you are also always there for me when I need you guys, and have helped me keep my feet on the ground during all the years and not to forget important family values despite us all living in different places. You guys all mean the world to me!
Of course I would also like to say a massive thanks to all my friends who have supported me throughout the years, and without whom I would have forgotten the other things in life that make life worth living. Matt, Tim, Borja, Dries Nicole, Karlien, Daniel, Chris, Dennis and Adam from my Utrecht days, I'm very happy we are still in touch and hope to remain so throughout the years, no matter where we all end up! Paul, Spencer, Arthur and Michael from the London era, as, although the 1.5 years were very short, they will always remain in my memory as an epic time. Likewise, thanks to Albert, Lidia, Greg and Damian, as you helped make Copenhagen the amazing experience it was. Not to forget that without you Damian, I would have never picked up the Python skills like I have now, so I am also extremely thankful for that. Additionally, thanks go out to Lars Juhl Jensen and his lab members for hosting me for a month back in 2009 to kickstart the computational biology side of my PhD, this was also an extremely fruitful time for me!
Last but most certainly not least, I can't express my gratitude to my partner Stina Rikke Jensen enough, who has supported me throughout most of my PhD and who understood when I needed to work late or during the weekend and who has always been there for me. I still remember the early days where you showed me how to count cells and where we had long discussions in the lab, which lead to far greater things than just collegiality. I really appreciate all the fun times we have together and reminding me to have fun even during extremely busy times, and I can't wait for all the adventures that await us in the future! :)

Articles included in the thesis

[1] Creixell P, Schoof EM, Erler JT and Linding R. *Navigating cancer network attractors for tumor-specific therapy*. **Nature Biotechnology** 30, 842 - 848. (2012)

[2] Schoof EM, Linding R. *Experimental and computational tools for analysis of signaling networks in primary cells*. **Current Protocols in Immunology**, Feb 4;104:11.11.1-11.11.23 (2014)

[3] Horn H*, Schoof EM*, Kim J*, Robin X, Miller ML, Diella F, Palma, A, Cesareni G, Linding R, Jensen LJ. *KinomeXplorer: a powerful integrated framework for kinome biology studies*. **Nature Methods**, Under Revision

[4] Schoof EM*, Creixell P*, Pasculescu A*, Kim J, Wesolowska-Andersen A, Gupta R, Linding R. *Uncovering hidden signaling networks by genome-specific proteomics*. Manuscript in preparation for submission to **Nature Methods**

[5] Schoof EM*, Cox TR*, Ferkinghoff-Borg J§, Longden J§, Pasculescu A, Creixell P, Kim J, Santini CC, Murray G, Erler JT, Linding R. *Targeting colorectal cancer metastasis using global, quantitative and integrative network biology*. Manuscript in preparation for submission to **Science**

Articles not included in the thesis

[1] Cox TR, Schoof EM, Rumney RMH, Agrawal A, Bird D, Ab Latif N, Evans HR, Huggins ID, Lang G, Linding R, Gartland A, Erler JT. *The hypoxic secretome reveals lysyl oxidase as a critical mediator of pre-metastatic bone lesions in breast cancer.* Manuscript in submission to **Nature**

[2] Djidja MC*, Chang J*, Hadjiprocopis A, Schmich F, Sinclair J, Mršnik M, Schoof EM, Barker HE, Linding R, Jørgensen C, Erler JT. *Identification of hypoxia-regulated proteins using MALDI-Mass Spectrometry Imaging combined with quantitative proteomics*. Manuscript in submission to **Journal of Proteome Research**

[3] Pasculescu A, Schoof EM*, Creixell P*, Zheng Y, Olhovsky M, Tian R, So J, Vanderlaan R, Pawson T, Linding R, Colwill K. *CoreFlow: A computational platform for integration, analysis and modeling of complex biological data*. **J Proteomics** Feb 3. pii: S1874-3919(14)00041-4 (2014)

[2] Zanivan S, Meves A, Behrendt K, Schoof EM, Neilson LJ, Cox J, Tang HR, Kalna G, van Ree JH, van Deursen JM, Trempus CS, Machesky LM, Linding R, Wickström SA, Fässler R, Mann M. *In Vivo SILAC-Based Proteomics Reveals Phosphoproteome Changes during Mouse Skin Carcinogenesis*. **Cell Reports** (2013)

[3] Creixell P, Schoof EM, Tan CSH and Linding R. *Mutational Properties of Amino Acid Residues - Implications for Evolvability of Phosphorylatable Residues*. **Phil. Trans. R. Soc. B**,vol. 367 no. 1602 2584 - 2593 (2012)

[4] Tan CSH, Schoof EM, Creixell P, Pasculescu A, Lim W. A., Pawson T, Bader, G. D., and Linding, R. *Response to Comment on Positive Selection of Tyrosine Loss in Metazoan Evolution*. **Science** 332, 6032, 917 - 920. (2011)

Abbreviations

MS	Mass Spectrometry
NGS	Next-Generation Sequencing
HCS	High-Content Screening
CRC	Colorectal Cancer

Chapter I

Introduction

1. Introduction

Ever since its conceptual discovery as early as 3000 BC in ancient Egypt[1], cancer has been eluding major scientific breakthroughs in terms of finding a cure. Despite large-scale efforts such as the 'War on Cancer'[2] in the 20th century, and estimated yearly research budgets as high as 14 billion Euros worldwide[3], the disease continues to have a devastating effect on people's health, with often lethal consequences. In this chapter, a breviloquent overview of cancer will be provided, focusing mainly on conceptual challenges that have arisen over the years and scientific approaches that have been developed in an attempt to tackle some of the many aspects of the disease. The process of metastasis, the spreading of the primary tumor through the body and growth in secondary organs, will be discussed, as this is the cause of over 90% of all cancer-related deaths[4]. The biology of colorectal cancer (CRC), the second leading cause of death from cancer among adults will be briefly explored. Additionally, the significance of moving from a pathway-centric interpretation of cellular signaling towards a more global signaling network approach will be highlighted. Finally, novel frameworks providing a platform from which to study several biological aspects will be introduced, with a special focus on Network Medicine and technologies fundamental to this, such as Mass Spectrometry, Next-Generation Sequencing, siRNA-based High Content Screening and computational modeling to allow the integration of different types of data. Together, it is envisioned that these genome-scale technologies will enable novel modeling approaches such as biological forecasting, which can predict cellular behavior based on in vitro and in vivo experimental data.

1.1 Cancer - a conceptual overview

Recently, Neanderthal fossils containing signs of cancerous growths have been described, which have been dated to more than 120,000 years ago[5], underlining how long the disease has had a detrimental effect on life. The exact definitions of cancer are continuously being updated as our knowledge of this disease increases. However, over the last few decades, major strides have been made to shed some light on the mechanisms underlying the complexity of the disease. Many of the significant accomplishments have given us a better understanding of what biological components make up a tumor, and the importance of each of these aspects in constituting potential biological processes suitable for therapeutic intervention.

Even though every cancer appears, at least at the genetic level, unique[6,7], it was proposed by Hanahan and Weinberg in as early as 2000, and re-iterated in 2011[8,9] that there are fundamental underlying traits that separate cancer cells from their healthy counterparts. While an extensive discussion of these hallmarks exceeds the scope of this chapter, it is important to consider the molecular mechanisms underlying these phenotypes. In other words, what is it that allows cancer cells to elude normal cellular control mechanisms and eventually grow and spread uncontrollably? As depicted in Figure 1, Hanahan and Weinberg defined six properties a cell must acquire in order to become malignant. Of these, the ability of cancer cells to sustain chronic proliferation is one of the most fundamental traits, and by overcoming the dependence on specific growth-promoting signals, cancer cells "become masters of their own destinies"[9]. This property can be considered the start of a cascade of tumor pathogenesis, where the five other hallmark traits can be acquired throughout tumor evolution, to eventually result in final metastatic and invasive disease. As will be explored more in-depth in Section 1.3.1, it is important to note that these attributes need not all occur in the same cell, as tumors often consist of heterogeneous populations of different cell types[10,11]. Within a tumor, the accumulation of mutations can occur via separate evolutionary paths, giving rise to genetically distinct sub-populations. Within these sub-populations, different cell types operate in a concerted way, where the phenotypic alterations of each cell type due to genomic alterations confers the required properties for e.g. tumor growth, sustenance and metastasis.

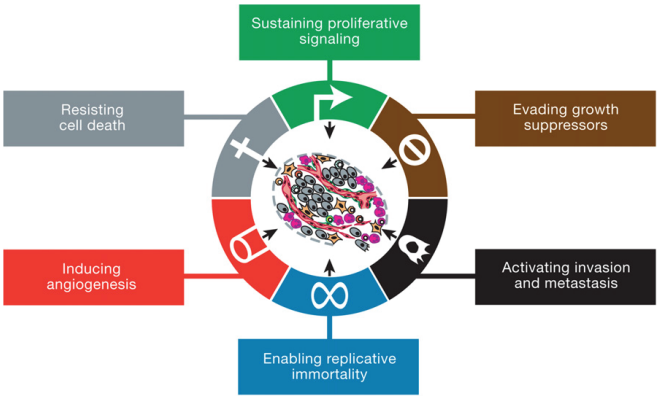


Figure 1 - The Hallmarks of Cancer, as proposed by Hanahan and Weinberg[8,9].

1.2 Effectors of Cancer

While extensive strides have been made in terms of defining biological characteristics that drive a tumor cell towards malignancy, there is still great debate about the precise underlying causes at each stage. As will be discussed throughout this thesis, several biological phenomena may be fundamental to all human diseases, ranging from mutations occurring in the genome of the diseased cells[6,10,12], through epigenetic regulation of DNA transcription[13,14], to dysregulated protein network dynamics[15-23]. Whilst each of these may contribute individually, it is likely that a combination of these aspects is what ultimately drives cancer and adds to the complexity of understanding a given disease phenotype. For example, it is known that more than 50% of human melanomas contain an activating mutation in the BRAF kinase gene, where the valine in amino acid position 600 is substituted with a glutamate (BRAF V600E), which, through structural effects, renders the kinase constitutively active[24]. This permanent kinase signaling results in a constitutive activation of the Raf to mitogen-activated protein (MAP) kinase pathway, significantly increasing the rate of mitogenesis of these cells[25]. While inhibition of this mutant-specific variant of BRAF often results in a temporary response to treatment in the clinic[26], resistance ultimately develops as the signaling networks normally utilizing the BRAF kinase pathway are re-wired to circumvent the BRAF inhibition[27,28]. Currently, the use of additional inhibitors to subsequently target the re-wired cellular signaling networks is being investigated, with some initial success[29]. This is a prime example of how mutations at the genome level exert their effect at the protein signaling level, and underlines the importance of investigating both genomic and proteomic aspects of a cell when trying to gain a thorough understanding of a given disease phenotype. In addition, while proteins are the cell's functional effectors and almost exclusively the targets of small-molecule inhibitors and antibodies, understanding the underlying genomes can help pin-point appropriate protein targets and aid in elucidating molecular mechanisms of a particular disease. Especially considering the fact that "no single gene defect 'causes' cancer"[12], understanding how proteins interact with one another in a signaling network through proteomic analysis can help prioritize which mutations may act in a coherent manner and should be further investigated in combination for therapeutic purposes [30]. While the complexity is extremely high at the genomic level, given the similar phenotypic states of cancer cells (increased proliferation, migration etc.), the complexity might be less profound at the protein signaling network level. In other words, while the mutational landscapes might differ greatly between tumors, their functional impact on the signaling networks within the cell might be less varied, thereby eliciting a similar phenotype. This underlines the need for assessing the impact of mutations at the protein level, which is explored further in Chapter 2.1[16], where we present a novel conceptual framework in which cancer signaling could be interrogated.

1.3 Tumor Complexity

Although the number of challenges in cancer research are virtually unlimited, there are three main concepts of tumor complexity that deserve special attention in the context of this thesis, as they highlight significant clinical challenges in the treatment of cancer.

1.3.1 Tumor Heterogeneity

The first aspect is that of tumor heterogeneity, which highlights the fact that within a tumor, not all cells are equal (see Figure 2). Different cellular clones often exist in unison, driven by independent evolutionary fates, each potentially displaying different phenotypes based on the genetic and proteomic landscapes they harbor. A leading theory, the cancer stem cell hypothesis, postulates that (some) cancers consist of a hierarchy of tumorigenic cancer

stem cells (CSCs) and their non-tumorigenic counterparts, and that the cancer stem cell sub-population is fundamental for driving tumor growth and disease progression[31-33]. Original work conducted by Furth and Kahn in 1937 established that, rather than a transmissible agent, a single cell from a mouse tumor was sufficient to induce cancer in a healthy recipient mouse[34]. Throughout the following decades, transplantation assays carried out by others revealed that the frequency of cells which can initiate cancer varies greatly among different solid tumors and leukemias, but generally requires a low number of cells (10^3 to 10^7 cells)[35-37]. Subsequently, a series of experiments conducted by Pierce and colleagues showed that within malignant teratocarcinomas (germ cell tumors), highly tumorigenic cells exist that, as single cells, have the capacity to differentiate into several differentiated, non-tumorigenic cell types[38]. With this, it was established that the maturation process of a tumor seems to occur in a similar manner to normal tissue development, with CSCs forming the foundation of a heterogeneous tumor population.

The concept of CSCs was thoroughly established through seminal work carried out by Dick and colleagues, where they showed that acute myeloid leukemia (AML) can be induced reliably in immuno-compromised mice, but only through implantation of cellular sub-populations displaying a $CD34^+CD38^-$ phenotype (as sorted through Fluorescence-activated cell sorting, FACS)[39]. Additionally, they were able to investigate the frequency of CSCs, which was found to be one in every million tumor cells in AML. Later publications showed similar results in tumors originating from breast cancer[40], brain cancer[41] and colon cancer[33], where a relatively low number of cells, displaying distinct cell-surface antigen profiles, were able to induce cancers in immunodeficient mice.

From a cancer treatment point of view, CSCs simultaneously pose both a challenge and an opportunity. On the one hand, CSCs have been attributed with increased resistance to treatment through, for example, quiescence mechanisms, increased expression of ATP-binding cassette transporter (ABC) drug pumps and anti-apoptotic proteins, and elevated resistance to DNA damage[42-45]. On the other hand, the notion of CSCs is clinically attractive, as it helps explain why patients with initially good treatment response and who have been declared 'cured', return to the clinic years later with recurring disease. In these cases, while the treatment successfully eradicated most of the tumor, a small number of CSCs may have survived and recolonized tumors both at the primary and metastatic sites. If one were able to target these CSC sub-populations specifically, this would potentially allow complete removal of a tumor. Early evidence of being able to specifically target CSCs is starting to emerge, with specific proteins having been related to the CSC phenotype, allowing directed therapeutic targeting of CSCs[46-48]. While additional research is required, these early results highlight the therapeutic potential that may be obtained. For further information about cancer stem cells, the reader is referred to previously published review articles[11,31,32].

1.3.2 Tumor Micro-Environment

The second aspect this section will briefly touch upon is the role of the microenvironment, as extensive evidence has demonstrated that it is not only the tumor cells themselves which decide the fate of pathogenesis, but rather that it is the complex interplay of the tumor cells with their surrounding microenvironment and other cell types contained therein (see Figure 2). In healthy tissue, the stroma (consisting of extracellular matrix (ECM) and other cells such as endothelial cells, fibroblasts etc.) is generally considered the supportive structure of a given tissue or organ, and maintains a natural barrier against tumorigenesis. When tumor cells appear however, this process initiates a cascade of changes, often resulting in the stroma becoming an environment that supports cancer progression. This involves, amongst others, the activation of fibroblasts and matrix remodeling, and additionally, micro-environmental stimuli such as hypoxia, acidity and growth factors will vary within a tumor, giving rise to additional heterogeneous cell populations, each adapted to their specific microenvironment. These micro-environmental factors are likely to affect both the signaling network states of tumor cells contained within, and may even select for specific DNA mutations that give certain sub-populations an evolutionary benefit depending on their surroundings.

A particular cell type that has been associated extensively with cancer progression, and thereby merits scientific investigation, is the cancer-associated fibroblast (CAF). Where normal fibroblasts typically have a suppressive effect on tumorigenesis[49], CAFs can significantly enhance tumor progression[50]. They distinguish themselves from normal fibroblasts by displaying increased proliferation levels, enhanced cytokine secretion, and elevated extracellular matrix production and increased contractility[51]. These differences result in extensive tissue remodeling,

predominantly around the tumor cells, driven extensively by e.g. elevated expression of matrix metalloproteases, pro-angiogenic factors, matrix cross-linkers such as lysyl oxidase (LOX) and transglutaminases, and growth factors[52,53]. The clinical relevance of CAFs has been established in several ways. For example, the abundance of CAFs has been shown to be predictive of prognosis for both breast and pancreatic cancer[54,55]. Additionally, elevated matrix metalloprotease levels, which are secreted both by tumor cells and CAFs, have also been linked with increased tumor aggressiveness and poor prognosis[56]. Furthermore, CAFs have been associated with emerging treatment resistance[57]. For example in BRAF(V600E) cell lines, co-cultured with CAFs, BRAF inhibition was overcome by secreted hepatocyte growth factor (HGF), which elicited increased phosphorylation of MET, a cognate receptor of BRAF[58,59]. This has also been observed in patients, where HGF expression levels positively correlated with reduced drug treatment response[58]. These studies highlight the importance of studying cancer cell signaling within the appropriate context, as the in vivo relevance of in vitro derived results will be limited.

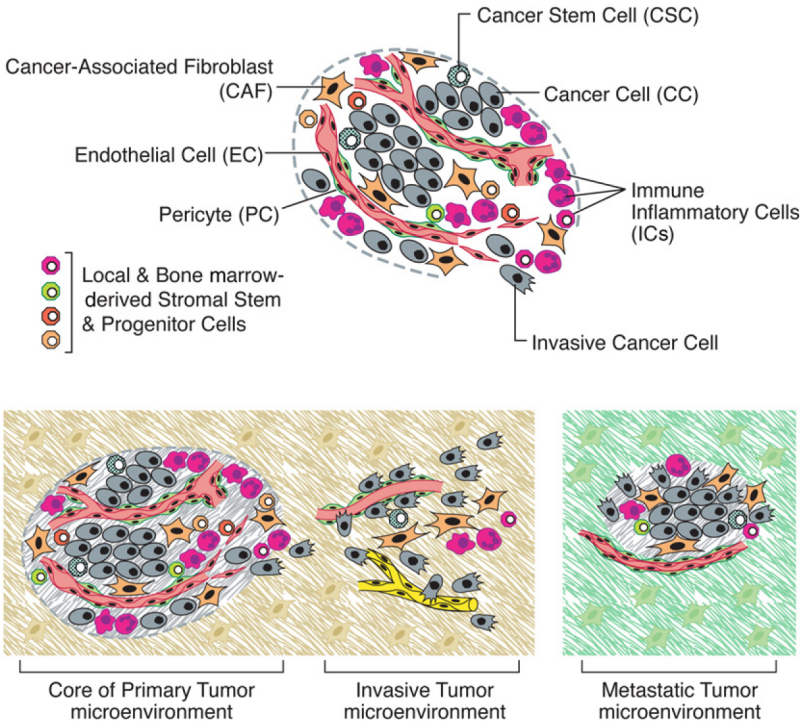


Figure 2 - Schematic representation of the key cellular components of the tumor microenvironment and cellular components contributing to tumor heterogeneity. In invasive and metastatic tumor microenvironment, there is an increase in number of invasive cells, contributing to the overall tumorigenesis. Taken from Hanahan & Weinberg[8,9].

To add to the complexity, the influence of the microenvironment on tumorigenesis is a bi-directional process. As was recently shown by Barker et al., lysyl oxidase-like 2 (LOXL2) is an enzyme secreted by tumor cells, which in turn activates the CAFs through integrin-mediated focal adhesion kinase activation. By blocking LOXL2, the authors reduced the activation of stromal host cells, and significantly decreased tumor cell invasion[60]. Similarly, Cox and colleagues investigated the effect of the microenvironment on the metastatic potential of a tumor[61]. They found that the expression of lysyl oxidase (LOX), an enzyme that catalyses the covalent crosslinking of Collagen I and which has been implicated in cell invasion and malignant progression[53,62], is "favourable to colonization and growth of metastasising tumor cells"[61]. By specifically blocking LOX activity using an antibody, they were able to reduce the extracellular matrix (ECM) modifications that normally have an enhancing effect on tumor cell survival and metastasis. As LOX is secreted from the 4T1 cell line used in the study, this elegantly highlights the complex bi-directional interplay of tumor cells and their surrounding microenvironment. Not only does the targeting of proteins secreted by the tumor cells seem like a viable therapeutic strategy though, as work by Luga and colleagues demonstrated that targeting CAF-secreted exosomes (using Cd81-specific siRNAs) has a beneficial effect on specifically suppressing lung metastases in a breast cancer model[63]. Combined, these results suggest that targeting both the tumor cells and surrounding stromal cells are attractive therapeutic strategies, and that more extensive investigation is merited to uncover additional molecular mechanisms driving tumorigenesis, both in tumor and their surrounding stromal cells.

In summary, these first two aspects underline the importance of moving away from viewing a tumor as a homogeneous population that can be treated as a whole, as it is evident that the complexity is much larger than initially assumed. Only by biologically characterizing the contribution of individual components that make up a

tumor, both separately and collectively and understanding how they can be therapeutically targeted, is it likely that cancer research will start fulfilling its promise of successful targeted therapies.

The third and final aspect of tumor complexity is the concept of cancer metastasis. Considering this is a large focus of Chapter 3.2, it deserves additional attention, which is why it will be discussed in more depth in the following section.

1.3.3 Cancer Metastasis

As was depicted above, the cascade of tumorigenesis often ends at the metastatic stage, at which point the tumor cells, after acquiring the necessary mutations and other phenotypic traits (e.g. invasiveness, enhance proliferative capacity etc.), have been able to colonize other organs within the body. This spread of cancer cells to other, often vital organs, is responsible for 90% of all cancer-related patient mortality, and, if targeted successfully, represents the hallmark with most therapeutic potential[4]. In this section, we will explore some of the underlying biological principles, challenges associated with characterizing metastatic cells and possible directions for gaining a better understanding of how to tackle metastatic disease.

The process of metastasis is often classified into a series of basic biological events, which are portrayed in Figure 3 (adapted from[64]). The steps involve 1) local invasion, 2) intravasation (entry into the bloodstream), 3) extravasation (exit from the bloodstream) and 4) colonization of the distant tissue, from which further metastases can also be spawned[65,66]. While each of these steps represents a potential opportunity for therapeutic intervention to varying degrees, it is important to consider the origin of these traits and how tumor cells obtain these capabilities, as this may lay the foundation for successful prevention; this will be further explored in the next section. As previously explained, the concept of cancers consisting of a heterogeneous population of tumor and stromal cells has now been widely accepted, each cell lineage contributing to the development process. Furthermore, research conducted in the last decade has suggested that several oncogenic events during cancer development may contribute to the evolution of tumors. Especially the ability to resist growth suppression and bypass DNA-damage-checkpoints are critical properties, as this allows for the generation of genomic instability[67-70] Particularly in the case of colorectal cancer (CRC) pioneering work by Vogelstein and colleagues was able to show that colorectal tumorigenesis largely relies on the linear accumulation of key mutations[71]. Through this, tumor cells are exposed to many evolutionary paths, some of which may give a specific sub-population e.g. a fitness advantage or the phenotypes required for metastatic progression. Additionally, the genetic and phenotypic diversity that thereby exists within a tumor may also explain why, despite millions of cells being shedded by a tumor into the blood stream every day, only a very small subset of these will successfully colonize distant tissue[72-74]. Part of the reason for this high attrition rate of distant organ colonization is the fact that healthy tissue generally provides an inhospitable environment for invading tumor cells, requiring a significant level of resilience to exist in metastatic tumor cells to overcome this. As originally postulated by Stephen Paget in 1889, the spread of metastasis is dependent on both the primary tumor (the ‘seed’) and the distant site (the ‘soil’)[75]. For an extensive review, the reader is referred to[66], but it is interesting to note that the metastatic progression does indeed seem to depend on both the primary and secondary tissue sites[4,76]. Thus, it becomes clear that it is beneficial for a tumor to contain specific sub-populations of tumor cells, all bearing different pheno- and genotypes, allowing the successful growth of the primary tumor and distant metastases in their respective environments. Pioneering work conducted by Fidler and colleagues in the 1970s demonstrated that within a tumor, rare clones exist that, through evolutionary processes, had acquired the necessary properties which allowed them to drive metastatic progression[72]. Later work revealed that highly metastatic sub-populations displayed a higher level of genetic mutability than their non-metastatic counterparts from the same tumor, in addition to biological heterogeneity being observed both within a single metastasis (‘intra-lesional’ heterogeneity) and among different metastases (‘inter-lesional’ heterogeneity), which supports the link between metastasis and genetic instability/evolution[66]. Overall, these results seem to suggest that cancer progression is a function of heterogeneous cell populations being driven to evolve through sequential environmental stimuli and pressures[4].

1.3.4 'Drivers' of Cancer Metastasis

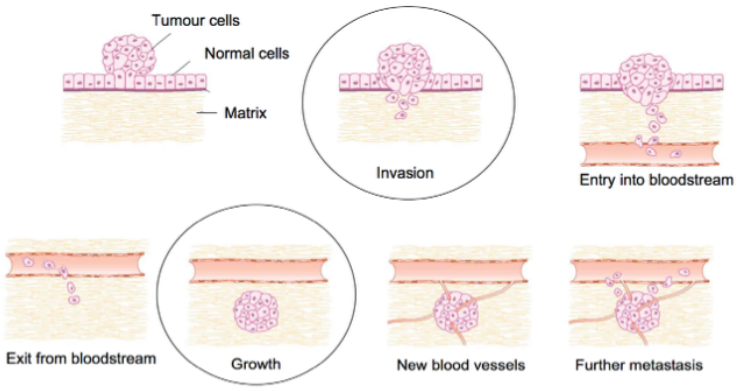


Figure 3 - The process of metastasis, depicting the different biological processes tumor cells must undergo to successfully colonize distant tissue sites. Adapted from Erler & Giaccia[64].

The dysregulated proliferation of tumor cells was proposed as one of the hallmarks of cancer[8], but the underlying processes of tumorigenesis towards a malignant metastatic phenotype is a combination of many intrinsic (e.g. genetic, epigenetic and protein dynamics) and extrinsic factors. A critical extrinsic factor which seems to specifically affect metastatic potential is, amongst those described in Section 1.3.2, the level of hypoxia within the tumor. Hypoxia is the lack of oxygen available to cells within a tumor, and already occurs in tumors only a few cubic millimeters in size. Therefore it is a common driver of tumor aggressiveness[77]. Hypoxia has been shown to promote the enrichment of sub-populations with a higher resistance to apoptosis, and the response involves the stabilization of hypoxia inducible factor-1 (HIF-1), which is a transcriptional complex promoting the expression of genes involved with angiogenesis, anaerobic metabolism and cell invasion and survival[78]. For example, HIF-1 induces the expression of CXCR4, a chemokine receptor, which has been demonstrated to promote renal cell carcinoma metastasis[79]. Additionally, the LOX enzyme mentioned above has been demonstrated to be regulated by HIF-1, increasing the rate of metastasis in a breast cancer mouse model[53]. Lastly, hypoxia has also been shown to increase the expression of Met kinase, which increases the rate of cell invasion mediated by HGF (which, as described above, is often secreted by CAFs)[80].
Other evolutionary driving forces promoting metastatic progression include reactive oxygen species of nitrogen and oxygen. These are generated both by infiltrating inflammatory and rapidly proliferating tumor cells, and have been demonstrated to up-regulate the expression of metastasis-facilitating genes, and contribute to the genomic instability of cancer cells[81]. Lastly, as demonstrated by Paszek and colleagues, the physical alterations during tumor development give rise to tensional forces, which may result in integrin-clustering. Integrins are widely used for mediating the attachment between a cell and its surroundings, and the tensional forces give rise to ERK and Rho-GTPase activation, thereby promoting tumor-cell proliferation and distorting tissue polarity[82]. This mechanical aspect has also been demonstrated by Cox and colleagues (briefly mentioned in Section 1.3.2), and subsequently by Baker et al., where increased tumor stiffness was linked to increased metastatic potential in a colorectal cancer model[61,83]. In both models, it was shown that Src kinase activity played a crucial role, and Focal Adhesion Kinase (FAK) was also implicated in the latter study, highlighting potential therapeutic targets to perturb the tumorigenic progression. Furthermore, these studies also elegantly highlight the possibility of studying intracellular molecular processes, which are occurring in response to extrinsic stimuli, to extract possible proteins that can be targeted with molecular intervention.

Despite a plethora of factors affecting the tumor aggressiveness at the primary site, other biological traits must be obtained as well, such as the capacity to initiate tumors at a remote site, alterations in cellular adhesion, resistance to extracellular death signals and the ability to migrate and invade. As an extensive discussion of these traits exceeds the scope of this chapter, the reader is kindly referred to the following publications:[4,76,84]. Nevertheless, it is important to consider the underlying molecular processes which facilitate these and the above-mentioned biological properties. Moreover, obtaining a more comprehensive understanding of which proteins seem to play a fundamental role in their development, and in particular which stages of development, is of critical importance, as they may each pose possible therapeutic targets. Kinases have been shown to be involved in many metastasis-promoting cell behaviors such as cell proliferation, migration, survival and invasion[60,61,83,85-91]], and thereby form attractive therapeutic targets[92-95]. In fact, it has recently been shown that kinases are the most frequently mutated proteins in

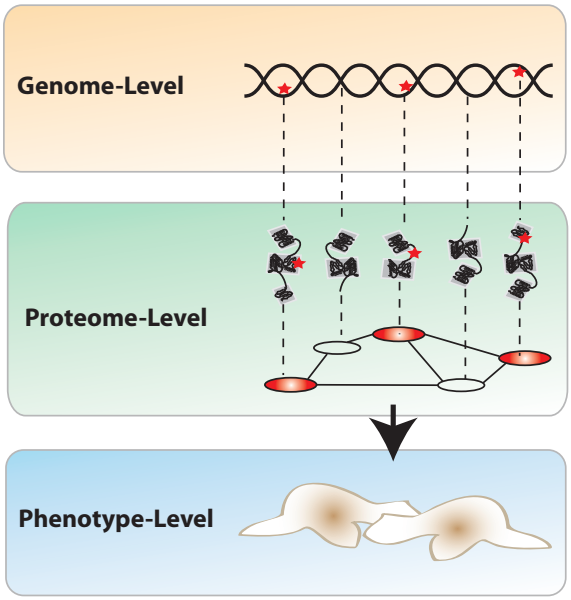


Figure 4 - The genotype-to-phenotype relationship, with protein signaling networks playing a critical role as a link between the two.

tumors, underlining the potential therapeutic implications they represent[96,97]. Given the diversification that tumor cells undergo throughout metastatic progression, and the ubiquitous involvement of kinase signaling in cellular signal processing[18,98-100], it is expected that kinase activities will be altered during this process as well. Therefore, obtaining a global overview of which kinases and other proteins are dysregulated during tumorigenesis is of great importance, also considering the plethora of inhibitors available for this group of proteins. As we explore extensively in Chapter 3.2, we have undertaken a genome-scale investigation and comparison of metastatic versus non-metastatic colorectal cancer cells, in an attempt to pinpoint specific proteins and kinases which may be fundamental to a metastatic phenotype, both at the genomic and protein network / kinase dynamics level. The technologies and conceptual frameworks which have been developed and improved in recent years, enabling this type of global characterization, will be briefly discussed in the final section of this chapter, after briefly discussing the biology of colorectal cancer below.

1.3.5 Colorectal Cancer

In Chapter 3.2, we describe a systems level approach to studying colorectal cancer metastasis in attempt to construct a metastasis-specific treatment strategy. Here, we will briefly explore the molecular basis of CRC to serve as a very brief summary of some of the milestones which have been achieved in the field. In the United States alone, every year 160,000 cases of CRC are diagnosed, giving rise to 57,000 mortalities annually and ranking the disease as second leading cause of cancer-related deaths[101]. The disease is generally described as originating from a benign adenomatous polyp, which through processes described above (genome instability, hypoxia etc.), can progress to a fully metastatic cancer. Patients that present with CRC at the clinic are classified into one of four stages, with Stage I and II tumors being confined to the colon, whereas Stage III and IV tumors have spread to the lymph nodes and further to other distant sites respectively[102]. Treatment of Stages I and II are generally curative through surgical excision, and 73% of Stage III tumors are curable by surgery combined with adjuvant chemotherapy[103]. For Stage IV tumors however, no cure is available and most patients succumb to the disease within 2 years[102]. As previously mentioned, genomic instability appears to be a fundamental requirement for the development of metastasis[71]. Chromosomal instability has been described as the most common type of genomic instability in CRC, which, by causing numerous changes in chromosomal copy number and structure, results in the physical loss of wild-type copies of tumor suppressor genes such as APC, TP53 and SMAD4[104,105]. Alternatively, subsets of CRC patients display inactivation of DNA mismatch repair genes such as MLH1, MSH2, TGFBR2, BAX and MYH, some of which are hereditary and others are acquired somatically[106-111]. Together, these genetic alterations give rise both to dysregulated signaling dynamics and increased susceptibility to further genetic modifications, which can both contribute to tumorigenesis.

Genetic modifications that are commonly associated with CRC can have both inactivating and activating effects on protein activity, with varying results on cellular signaling. For example, the activation of Wnt signaling, regarded as the first initiating event in CRC, is caused by a mutation in APC. APC is part of the β-catenin degradation complex (together with GSK3, axin and casein kinase 1[112]), which normally degrades β-catenin and prevents its nuclear localization where it binds to nuclear partners and creates a transcription factor leading to cellular activation[113]. After mutation of APC, entry of β-catenin into the nucleus cannot be prevented, thereby constitutively activating Wnt

signaling. The resulting phenotypes (hyperproliferation, perturbed differentiation and migration[114]) have furthermore been demonstrated to be dependent on c-Myc, a transcription factor[115,116]. APC mutations have been described in both hereditary CRC (e.g. familial adenomatous polyposis, FAP) and somatic tumors, where both copies of APC are inactivated. Additionally, mutations in β-catenin have also been described in APC wild-type tumors, thereby rendering the protein resistant to the β-catenin degradation complex[113,117,118]. Other common tumor suppressor inactivations that have been described in CRC are TP53 and TGF-β, which are described as the second and third key genetic alteration for disease development respectively. Both TP53 alleles are inactivated in most tumors, which is often the result of a missense mutation which inactivates the transcriptional activity in one allele, and a 17p chromosomal deletion which results in the deletion of the second allele[119-121]. The loss of TP53 activity leads to a lack of cell-cycle arrest and a cell-death checkpoint, and intriguingly, the inactivation of TP53 is often related to the transition of the CRC from an adenocarcinoma to an invasive carcinoma[122,123]. TGFBR2 inactivating mutations are observed in approximately 30% of CRCs, and have a mainly somatic origin, resulting in distinctive frameshift mutations within its parent gene[124]. In approximately half of the CRCs, these mutations affect the kinase activity of TGFBR2, but mutations affecting targets downstream, such as SMAD2, SMAD3 or SMAD4 have also commonly been described, resulting in attenuated transcriptional control[97,124-129]. Mutations resulting in constitutive activation of specific signaling molecules have also been described, such as RAS and BRAF, which leads to hyper-activated MAPK signaling[24,130-132]. Additionally, mutations affecting PI3K and causing constitutive activation have also been described in approximately 30% of all CRC patients, resulting in increased AKT and PAK4 signaling and subsequent cellular proliferation[133,134]. On average, Stage IV CRC tumors contain 76 mutated genes, but the large patient-to-patient variation has highlighted the significant genetic heterogeneity that exists within CRC[105]. This greatly hampers the functional interpretation of these mutations, and this aspect is explored in more detail in Chapter 2.1.

Currently, treatment options for CRC is predominantly centered around chemotherapy, with the inherent toxic side-effects, underlining the critical need for more targeted approaches. Initially, treatment consisted of 5-Fluoroacil (5-FU, a thymidylate synthase inhibitor) in combination with leucovorin (LV, a 5-FU enhancing agent)[135,136]. Irinocetan, a topoisomerase I inhibitor was also investigated for treatment efficacy, as was Oxaliplatin, a DNA synthesis inhibitor. Subsequently, two main treatment regiments were developed, termed FOLFIRI (LV/5-FU/irinocetan) and FOLFOX (LV/5-FU/oxaliplatin), which are both currently used in the clinic[137-141]. In patients with Stage IV CRC, these treatments prolong survival for approximately 6-9 months, and differ mainly in their toxicity profiles, with FOLFOX affecting the nervous system whereas FOLFIRI affects the gastrointestinal system[142]. Some early attempts at targeted therapy have been investigated, with monoclonal antibodies against VEGF-A (bevacizumab) and EGFR (cetuximab and panitumumab) having been investigated, but all with very limited efficacy[143-147]. Taken together, these results highlight the inherent complexity of studying and treating CRC, and highlight a critical need for more efficient therapies to be developed. To this end, we set out to gain a better understanding of the molecular landscapes that may underlie metastatic CRC, which is explored in depth in Chapter 3.2.

2.0 Network Biology in Cancer

2.1 Principles of Network Biology

The last decade has seen a considerable shift of moving away from a ‘one drug, one target’ approach[148], as, especially in the case of complex diseases such as cancer, deploying highly specific compounds targeting a single molecular entity has led to extensive treatment failure[149-151]. Traditionally, the single-target approach focused primarily on individual ‘pathways’ that were found to have been dysregulated in disease. However, it has become increasingly clear that the proteins making up a specific pathway have many interactions outside their respective pathways, and it is therefore imperative to gain a better understanding of how individual pathways are assembled into higher order networks[22]. This aspect is further highlighted by the genetic complexity and heterogeneity of tumors, as the mutations spanning a large multitude of proteins will have a different effect on the protein dynamics,

yet still result in similar disease phenotypes. In other words, several genotypes can give rise to the same phenotype, while a single genotype can also give rise to several phenotypes, depending on the cellular context. These 'genotype-to-phenotype' relationships are one of the fundamental aspects the field of network biology is aiming to resolve, as, being the cellular effectors, protein signaling plays a critical role in defining the link between genotype and phenotype (see Figure 4)[15,16,22,23,152,153]. Rather than characterizing signaling pathways in isolation, it is therefore of vital importance to comprehensively measure genome-wide protein dynamics within the cell, as this allows the characterization of their global network structure and dynamics and how genomic information is propagated towards cellular phenotype[21,22,154-156]. Additionally, it has been proposed that it is not only individual protein species that make up a signaling network that can be therapeutically targeted; instead the structure and dynamics of a particular signaling network poses a powerful drug target by modulating the information flow through such a network[23]. This was further exemplified by work demonstrating that cells utilize the same protein network components for different cellular responses, and instead alter the network utilization of these to elicit a specific response. This highlights the importance of identifying how cellular networks are deployed specifically in a given disease condition, in order to be able to target the dynamics through disease-specific perturbations[150,157,158]. Furthermore, by measuring global protein dynamics within the cell, it enables a more comprehensive interrogation of how the effect of a perturbation affects the signaling network as a whole. The importance of this was first elegantly demonstrated through seminal work by Janes and colleagues in 2005[159]. The authors investigated whether the phosphorylation state of Jun-activate kinase (JNK) was indicative of pro-apoptotic or anti-apoptotic activity, and concluded that the phosphorylation state alone did not suffice to characterize the phenotypic effect of JNK. Instead, they were able to show that activation of JNK could lead to both apoptosis and proliferation, and that the outcome was dependent on the prior signaling network state at the time of activation. This landmark study conclusively proved that studying the activity of a single protein in isolation is not sufficient to accurately characterize its cellular role, and rather, that the contextual network it is operating within also needs to be interrogated in order to obtain a correct readout. Similarly, utilizing this contextual information allows the characterization of the signaling network rewiring that occurs in response to a perturbation, in order to target these altered network states with an additional perturbation. The power of this type of approach was demonstrated by Lee and colleagues in 2012, where they were able to greatly increase the level op apoptosis in breast cancer cells by designing a time-staggered combination treatment in a data-driven way[89]. More information about these principles can be obtained in Chapter 2.1 of this thesis[16].

In addition to protein dynamics (i.e. expression levels) altering cellular phenotype, extensive research has demonstrated that post-translational modifications (PTMs), such as phosphorylation, also play a crucial role in controlling cellular behavior[18,91,160,161]. The importance of phosphorylation is further supported by the fact that by the year 2010, 149 inhibitors targeting 42 kinases (the 'writers' of phosphorylation modifications) have been subjected to clinical testing, highlighting the therapeutic potential of interfering with PTM signaling[93]. A significant challenge with deciphering phosphorylation based signaling however, is the derivation of kinase-substrate interactions. This is explored in depth in Chapter 2.3, but due to the highly transient nature of kinase-substrate interactions, it is inherently challenging to experimentally determine which kinase(s) is responsible for a given phosphorylation site. Therefore, generally, a combination of experimental and computational analysis is required, to which end algorithms such as NetPhorest and NetworKIN have been developed[98,162,163]. These algorithms allow, based on both experimentally determined linear motif preferences of kinases and their network contextual information, to predict which kinases are likely candidates for experimentally observed phosphorylation sites (see Figure 5). The linear motifs (specific sequence preference around the phosphosite) are predicted by NetPhorest, and STRING (a protein-protein association database) adds the network contextual information on whether or not the kinase and substrate protein (containing the phosphorylation site) are known/predicted to interact in vivo. Combined, they are amongst the most comprehensive and accurate algorithms currently available compared to previously published frameworks[164-167]. The way these algorithms can be applied in order to study cellular signaling is extensively covered in Chapter 2.2, where we describe both how to generate global quantitative phospho-proteomics data and the modeling thereof using the novel integrated KinomeXplorer framework to pin-point key kinases which may be fundamental to a given cellular phenotype.

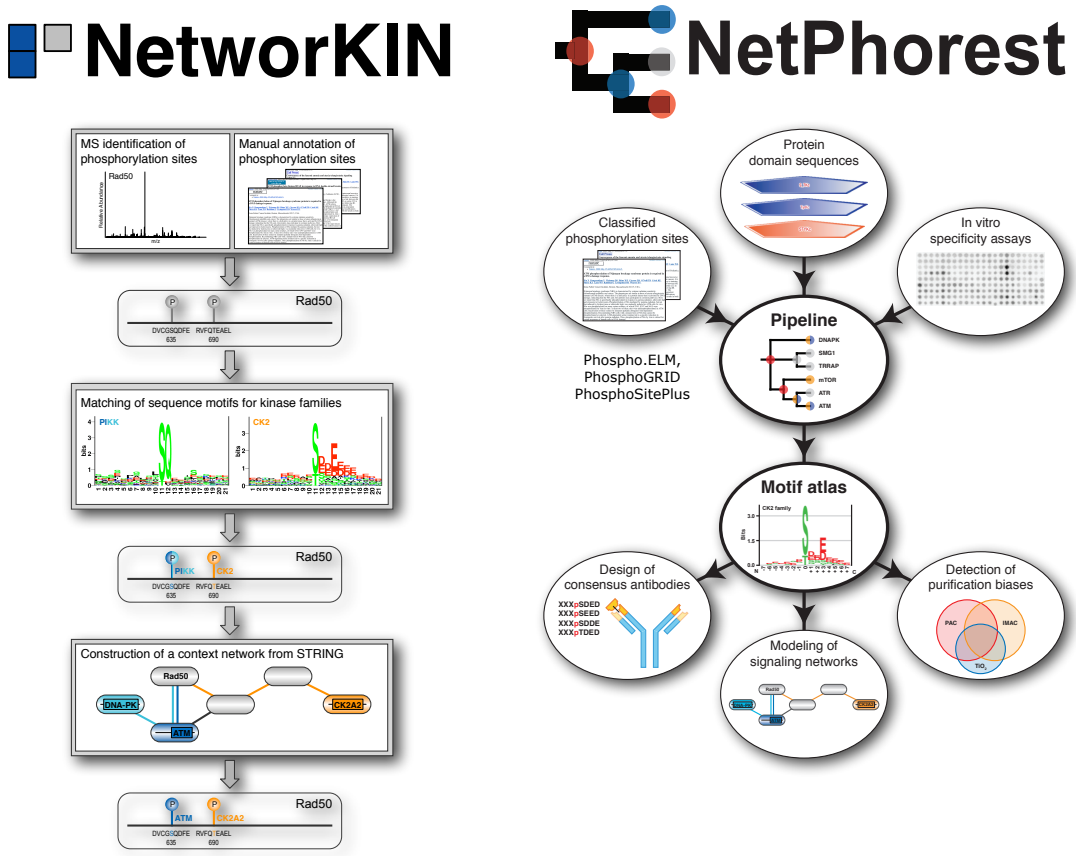


Figure 5 - Overview of the NetworKIN and NetPhorest workflows. By integrating motif-level predictions originating from Netphorest with protein contextual information from STRING, NetworKIN allows prediction of in vivo kinase-substrate interactions based on experimentally observed phosphorylation sites[98,162,163].

2.2 Tools to Enable Network Biology

In order to facilitate comprehensive sampling of the genotype-to-phenotype relationship, it is imperative that we can measure as many of the underlying properties (e.g. genome, epigenome, metabolome, proteome and cellular phenotypes) as possible, in a global, unbiased manner. Subsequently, these different biological aspects can be integrated through an integrative model in order to establish causal relations that may exist between them. From a practical point of view, this generally means that large experimental and computational infrastructures are required, to allow the generation and analysis of the required large-scale datasets in a timely fashion. We will now briefly explore some of these technologies in the light of trying to decipher cellular signaling in cancer using a Network Biology approach.

In order to systematically assess the genomic landscapes within a tumor, next-generation sequencing (NGS) of DNA is the method of choice, as it allows the determination of sequence variants either within the complete genome, or only the protein-coding exome. This type of analysis results in a complete view of the genomic landscape that exists within a sample, and allows one to begin inferring which genes might play a role in a particular disease phenotype without any prior knowledge[7,168]. With the significant decrease in costs associated with having a genome sequenced over the last few years, more and more laboratories obtain the capability of conducting such experiments, which has led to an explosion in publications analyzing diseased genomes[169-173]. Nevertheless, a direct therapeutic impact that was expected to arise from these analyses has been rather limited[30], and some reasons for this are explored in Chapter 2.1. One fundamental shortcoming of using genomic data alone is the fact that it does not allow

interpretation of how these mutations affect, or are utilized by, the cellular signaling networks. In other words, additional experiments are required for deducing which mutations have a functional impact, and are thereby functionally related to the disease phenotype[174]. For example, detecting a mutation at the DNA level does not guarantee that the mutation actually is expressed at the protein level, where it could exert an effect. Additionally, it may not be the mutations themselves which have a phenotypic effect, but rather, in case of kinases being affected by a mutation, the dysregulated phosphorylation-based signaling or network-rewiring that is caused by the mutation[16]. This underlines the need for actually being able to assess the protein dynamics of both the mutant proteins and their non-mutated network partners, and the dynamic phosphorylation networks which may be altered due to the genomic alterations. Additionally, the functional effect of a given mutation may only be apparent under specific cellular conditions, underlining the need for studying cellular behavior in response to several stimuli (e.g. growth factor stimulation, starvation, hypoxia or stromal co-cultures).

Recent developments in the field of Mass Spectrometry (MS) have led to major improvements of what portion of the expressed proteome can be measured in a single experiment. Currently, it is possible to simultaneously detect and quantify the protein levels of almost the complete proteome and several tens of thousands of phosphorylation events[175-181]. Quantitative technologies such as SILAC or dimethyl labeling allow the direct comparison of specific proteins or e.g. phosphorylation sites between one or several proteomes, in order to determine the dynamic regulation thereof[182,183]. Their principle is based on utilizing stable isotopes to introduce an small mass difference into the peptides that exist within a given sample, which allows the mixing of samples very early on in the sample preparation. This mass difference is large enough for a high resolution mass spectrometer to determine the sample of origin, and through direct comparison of the peptide intensities originating from the respective samples, enables direct quantitation of these peptides. As the samples were mixed in the early stage of sample preparation, any difference in intensity will likely be due to a biological effect rather than a technical artifact. If used in combination with appropriate experimental design, this type of quantitative analysis enables the in-depth characterization of the dynamic regulation of protein signaling after a given perturbation such as inhibitor treatment, growth factor stimulation, mutations or altered growth conditions (e.g. hypoxia). This information can subsequently be used to determine the signaling networks that are fundamental to a given biological observation. We deployed the SILAC methodology in Chapter 3.2 to study the dynamic proteome differences between metastatic and non-metastatic cells, and describe how it can also be used in the characterization of patient samples[184].
Additionally, as is described in-depth in Chapter 3.1, the use of genome-specific, rather than reference genome search databases for determining the proteins and phosphorylation sites present in a given sample, allows the accurate monitoring of the protein dynamics of specific mutations present in sample. Being able to quantitatively assess the information layer which exists between the genotype and phenotype has a distinct advantage, especially when considering that all small molecule inhibitors and antibody-based therapeutics used in the clinic today actually target proteins rather than genes. As was described above, these protein data need to be accurately modeled in order to establish potential signaling network models, but the near unbiased nature of MS experiments allows this to be done in a data-driven way with minimal *a priori* knowledge. Nevertheless, MS data only provides a "snapshot" of the cellular proteome at the time of the experiment, so experiments need to be designed appropriately to accurately assess the signaling networks fundamental to a particular biological effect.

In order to more directly characterize the role of key proteins in a given cellular phenotype, the use of genome-scale RNA interference (RNAi) libraries in combination with High Content Screening (HCS) has been demonstrated in several species[185-188]. By perturbing the expression of a specific protein, and measuring an appropriate phenotypic readout (e.g. proliferation or cell migration), it is possible to globally assess and elucidate proteins which seem fundamental for a given biological state. This approach has been very powerful not only at the discovery phase of investigating potential therapeutic targets, but also at the stage of validating proteins determined by e.g. MS or NGS to be involved in a given disease phenotype[189,190]. One of the strengths that is fundamental to the success of RNAi-based screening is the fact that they can be used as a highly specific substitute for often unavailable inhibitors. A caveat is that the link between RNAi and inhibitors is not always linear however, and requires careful validation[191].

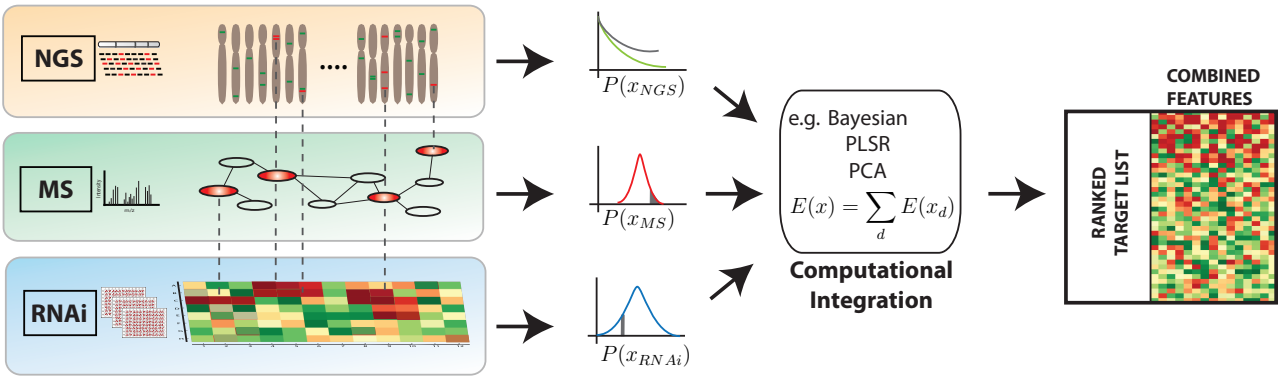


Figure 6 - Conceptual workflow for conducting integrative network biology experiments. By interrogating a biological system from the genome (NGS), global proteome (MS) and phenotype (RNAi) perspective, we can systematically assess the role of particular gene mutations and proteins in a given disease. Through computational integration of these datasets in an un-biased manner, potential targets can be ranked based on the amount of evidence that is highlighting their importance.

Nevertheless, by systematically knocking down genes in, for example, a kinome-wide or genome-wide fashion, a global assessment of the model system can be undertaken. Additionally, this can be done in a combinatory fashion, where multiple targets can be knocked down simultaneously to monitor synthetically lethal effects where a phenotypic readout is only obtained by the simultaneous knockdown of 2 or more genes[192-195]. Besides providing a more direct readout of the role of a particular protein in a given disease phenotype, it may also highlight proteins that were not detected by MS or genes that do not harbor a mutation, providing a complementary approach.

By subsequently integrating the knowledge gained from the different platforms (MS, NGS and RNAi screening), several benefits can be obtained: 1) missing information in one dataset can be complemented by that from another, and 2), the different data types can act as validation sets for one another, and 3) a comprehensive overview of how genomic information is propagated through the proteome to elicit a specific phenotype can be obtained. For example, a hit originating from an RNAi-screen is more likely to be a successful clinical target if MS evidence also suggests the protein-level (or phosphorylation dynamics) to be increased or if NGS has revealed there to be a mutation hitting the gene. In the case of point 1), if for example a given protein has not been observed by MS to be dysregulated in the diseased state, but the RNAi screen does confidently highlight the knockdown to result in a phenotype, it is likely to be an interesting candidate. Therefore, it is not only the genes and proteins which have been determined by all of the experimental methods to be of therapeutic interest, it is also the complementary nature of the technological platforms which allows for a more comprehensive interrogation of the biological model system. The lack of complete overlap between the different datasets does require an un-biased computational method for the data integration, and we will further explore a method we developed to this end in Chapter 3.2. Many different approaches for complex data modeling have been developed over the years, each with a unique, specific goal in mind[196-200]. Which approach is most suitable for which type of data depends on the precise biological question of interest, but most frameworks such as Partial Least Squares Regression (PLSR), Principal Component Analysis (PCA) or Artificial Neural Networks (ANNs) attempt to link a phenotypic readout (e.g. metastatic or not?) to observed experimental data (e.g. protein/phosphosite expression levels) to derive which proteins may be relevant for follow-up studies. In our integrative approach, we are assessing the significance of a particular observation within its own dataset (e.g. how much is a specific protein / phosphosite regulated in comparison with all the other proteins we were able to quantify or how much is a particular kinase knockdown affecting cell numbers compared to all other kinases tested), and are therefore quantifying the position of each observation within the general distribution for that dataset. This enables us to classify the level of “surprise” (termed “energy”) of an observation within the type of experimental analysis, and combine it with the “energy” levels for that protein of interest detected by the other experimental analyses that were conducted. Ultimately, this results in an integrated “energy” value for each protein, which is indicative of its potential involvement of the biological phenotype under investigation.

In conclusion, we advocate the in-depth interrogation of biological samples from several technological angles and biological perspectives, in order to accurately establish genotype-to-phenotype relationships (see Figure 6). Ideally, where possible, this should be conducted across several time-points and under different biological conditions, to interrogate the biological system as comprehensively as possible, as certain phenotypes or protein dynamics may only be revealed under highly specific biological conditions. Importantly, by taking a global, un-biased approach to interrogate a biological system, one allows the data to highlight which proteins and/or mutations are potentially fundamental to a specific disease phenotype. This requires the implementation of careful experimental design (tailored to the specific biological question of interest), strict Standard Operating Procedures (SOPs) for stable and reproducible sample preparation conditions and continuous quality control to ensure data is generated consistently over the generally extensive period of time it takes to collect the data. The practical application of this integrative approach is demonstrated in Chapter 3.2 of this thesis, which, to the best of our knowledge, is the first comprehensive network biology-focused study of cancer metastasis. In line with the above-mentioned role of tumor sub-populations and the micro-environment, ideally this would be done in a cell-type-specific manner (e.g. tumor cells, CAFs, CSCs and other sub-populations), as this will partially deconvolute the complex interplay of the different cell types, and potentially allow for time-staggered, cell-type-specific combination therapies to be developed, thereby positively contributing to the “war on cancer”.

Chapter II

Part I

Navigating Cancer Network Attractors for Tumor-Specific Therapy

Navigating cancer network attractors for tumor-specific therapy

Pau Creixell[1], Erwin M Schoof[1], Janine T Erler[2] & Rune Linding[1]

Cells employ highly dynamic signaling networks to drive biological decision processes. Perturbations to these signaling networks may attract cells to new malignant signaling and phenotypic states, termed cancer network attractors, that result in cancer development. As different cancer cells reach these malignant states by accumulating different molecular alterations, uncovering these mechanisms represents a grand challenge in cancer biology. Addressing this challenge will require new systems-based strategies that capture the intrinsic properties of cancer signaling networks and provide deeper understanding of the processes by which genetic lesions perturb these networks and lead to disease phenotypes. Network biology will help circumvent fundamental obstacles in cancer treatment, such as drug resistance and metastasis, empowering personalized and tumor-specific cancer therapies.

Cells are constantly computing decisions based on the integration of different cues that reach them at various times. In contrast to single-cell organisms, in multicellular organisms, cellular decisions should, ultimately, benefit the organism as a whole, even if that implies that an individual cell will have to decide to commit suicide. In line with this unique feature, signaling networks have evolved during multicellular evolution to allow cells to integrate cues and make decisions that ensure cooperative behavior between them. By hijacking these mechanisms, cancer cells escape cooperative rules and transition from a game governed by Nash equilibria[1,2] between all cells into a new scenario where cancer cells decide their behavior purely based on their own benefit, or as phrased by Hanahan and Weinberg[3], "become masters of their own destinies." Given the central role played by signaling networks in the integration of cues to compute any cellular responses, we argue that cancer is not simply a disease with a genetic basis, but is one ultimately driven by perturbations at the signaling network level, and that both the 'cue-signal-response' rules of cellular decision-making and the switch in strategy from cooperative to selfish are major, hitherto understudied, hallmarks of cancer[3,4].

In this article, we dissect the strategies cancer cells use to become 'selfish' and drive disease. We first review how genetic lesions can lead to altered protein function, which can result in changes to the structure and dynamics of signaling networks and ultimately cellular phenotype. Next, we describe five general properties of cancer signaling networks (**Fig. 1**) and define five challenges in cancer network biology and propose strategies to overcome them (**Fig. 2**). By meeting these challenges, network biology may fundamentally advance not only basic biology but also patient treatment. Finally, we describe how a combination of relatively new technologies could become a potent cocktail for the discovery of network drugs, and we discuss the practical implementation of personalized and tumor-specific cancer therapy.

From genomic lesions to functional network perturbations

Tumor cells often harbor hundreds to thousands of genetic lesions. But based on the observation that some of these genetic lesions are repeatedly observed in several cancers (e.g., *BRAF V600E*, present in >50% of all malignant melanomas[5]), it has been hypothesized that only a few genetic lesions are causally implicated in cancer development ('drivers'), whereas the majority have no functional consequences ('passengers')[6].

Although this classification has had some use in identifying mutations that are highly prevalent, it is now apparent that a tumor is not, under any circumstances, a static and uniform population of malignant cells. Rather, it is a dynamic ensemble of subpopulations with different abnormalities undergoing molecular evolution[7–9]. Two fundamental principles of cancer signaling networks can explain why a binary driver/passenger classification may be too simplistic to accommodate the complex dynamic nature of tumors. First, different tumors can develop similar phenotypes by acquiring mutations in different proteins[10], in what we term analogous mutations (**Fig. 1a**). Second, it has been shown that two different mutations not capable of causally driving cancer by themselves are able to do so when they appear in combination within the same cells or even within two neighboring cells[11], in what could be described as two passengers becoming drivers or, as we refer to them, synthetic oncogenes (**Fig. 1b**). Thus, patient-to-patient heterogeneity can be driven by the presence of different mutations in the same or in different proteins that lead to a similar signaling state and phenotypic outcome.

Altogether, the intrinsic heterogeneity of tumors makes it a pressing challenge for cancer network biologists to develop tools to identify the extent to which combinations of cancer mutations affect protein function and cellular and phenotypic states (**Fig. 2a**,**b**). Even though several such tools have been developed (reviewed in ref. 12), existing methods are mainly based on protein structure and/or sequence conservation. This is at odds with recent findings that show that cancer mutations tend not to cluster on the most conserved protein regions. In kinases, for example, mutations typically hit the kinase activation segment, a functional, yet largely nonconserved protein region[13].

[1]Cellular Signal Integration Group (C-SIG), Center for Biological Sequence Analysis (CBS), Department of Systems Biology, Technical University of Denmark (DTU), Lyngby, Denmark. [2]Biotech Research & Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark. Correspondence should be addressed to J.T.E. (janine.erler@bric.ku.dk) or R.L. (linding@cbs.dtu.dk).

Published online 10 September 2012; doi:10.1038/nbt.2345

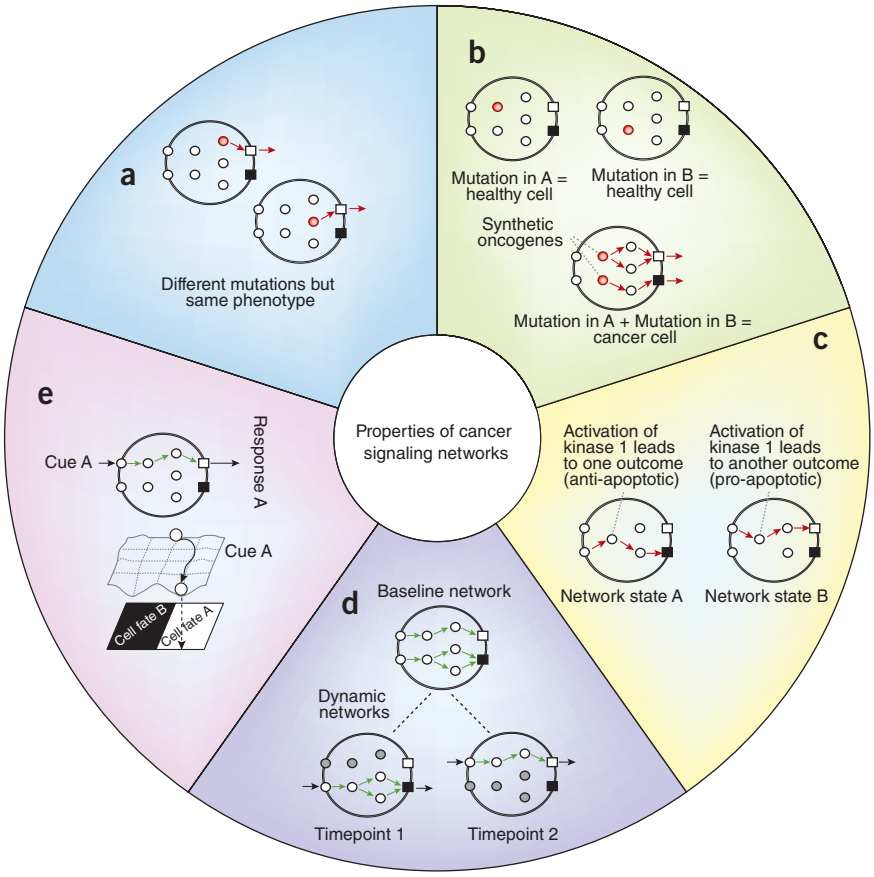


Figure 1 Properties of cancer signaling networks. (**a**) Analogous mutations. Two different tumors may achieve the same signaling and phenotypic outcome with two different mutations (**b**) Synthetic oncogenes. Mutations that are not oncogenic on their own can cooperate when appearing together to drive tumor formation[11]; by analogy to synthetic lethality, we call the genes harboring cooperative mutations, synthetic oncogenes. (**c**) Multivariate nature of signaling networks. The response of a cell to a specific cue depends on, and can only be predicted by taking into account, the state of the cellular signaling networks[25]. This dependency, known as the multivariate nature of signaling networks, is often neglected when classifying mutations and genes as oncogenes or tumor suppressors and cancer drivers or passengers. (**d**) Dynamic networks. Although signaling networks are often represented as static, it is clear that they are highly dynamic entities. Given that the role of signaling networks in computing cellular responses is highly dependent on it, and that cancer mutations will perturb it, this dynamic nature is a critical property of cancer signaling networks. (**e**) Signaling network landscapes. The different states that a signaling network occupies can be represented as a landscape (with stable steady states or attractors represented as valleys and unstable steady states represented as hills), where the cell constantly gets pushed by signaling cues[31,32,39,40]. These states drive cellular and disease phenotypes and represent network drug targets.

Because cancer cells would obtain the greatest fitness advantage from mutations that target the most-functional residues, we reason that a better understanding of the functionality of protein residues would allow more accurate predictions of the consequences of cancer mutations. Functional residues have been defined as those residues required for a protein to perform its molecular function(s), in the sense that they cannot be freely changed without directly affecting the role(s) of the protein[14]. Here we extend this definition to include a more fine-grained and precise definition of protein function as an ensemble of protein features that together describe the different functional capabilities of proteins (e.g., ATP binding, substrate specificity, protein activation or phospho-tyrosine binding). This new definition would not only adapt well to current studies of sequence-function associations[15,16], but also lead to a better description of the effects of a mutation affecting such residues (**Fig. 2a**,**b**).

An insightful example of how to explore this sequence-function relationship in protein domains was carried out by researchers in the Ranganathan and Yaffe laboratories who, using methods from statistical mechanics, generated synthetic WW domains *de novo* that maintained fold and function[17,18]. Further supporting a complex sequence-function relationship, additional studies from the Ranganathan laboratory demonstrated that, in addition to protein architecture described as combinations of modules such as globular domains and linear motifs[19–21], protein domains themselves often have well-defined sectors formed by sparse networks of residues often linking spatially distant regions that contribute cooperatively but unequally to its function[22,23]. Although some targeted studies analyzing several cancer mutations in a single kinase have been conducted[24], similar approaches to those used for WW domains should be pursued to generate high-throughput experimental studies of cancer mutations in the context of signaling networks. These would help gain a better understanding of which amino acid residues can be changed freely without affecting the protein and network function and, most importantly, which cannot.

From network perturbations to cellular phenotypes

The characterization of cellular signaling processes has largely focused on identifying the function of individual genes and proteins. A notable exception is a landmark study[25] on the context dependence of the Jun-activated kinase (JNK) in apoptosis. Before this work, paradoxical results suggested that JNK had a pro-apoptotic function[26], an anti-apoptotic function[27] or even a lack of involvement in apoptosis[28]. The systematic approach undertaken by Janes *et al.*[25] revealed that the phosphorylation status of JNK (and thus its catalytic activity) was not sufficient to determine apoptotic commitment; instead, activation of JNK could lead to both apoptosis and proliferation depending on the cellular signaling network state at the time of activation. Thus, this work demonstrated that a protein's cellular role is not a static property but rather can only be defined dynamically—that is, its role depends on the context of the network it is operating within. Similar context dependencies have been confirmed for other kinases, such as Erk and MK2. Because of this, which is referred to as the multivariate property of signaling networks (**Fig. 1c**), we suggest that it is essential to study cellular context at the systems level.

Although these multivariate molecular networks seem to have evolved a complex structure that makes them robust against deletion of a few proteins[29], they are highly dynamic. Thus, a more accurate description of signaling networks should take into account the fact that a single static network does not exist unchanged over time. Instead, a cell contains a dynamic ensemble of networks whose different permutations are manifested in the cell depending on the different cues the cell is presented

with (**Fig. 1d**). This dynamic nature of signaling networks could, at least in part, explain why all mutant proteins do not seem to be expressed at a given point in time[30], if a substantial part of the proteome is so dynamic that it is expressed only when the cell senses a specific cue.

Moreover, according to a general principle of complex systems introduced in the 1980s[31,32], dynamic cellular networks can only exist in a finite number of states, owing to the constraints that interactions between nodes impose on one another. These network states can be represented as landscapes, where most-probable and least-probable states are represented as valleys and mountains, respectively (**Fig. 1e**). Cells are continuously exploring this landscape and are pushed from one state to another by different environmental or intracellular cues.

Implications for cancer research

The multivariate nature of signaling networks has profound implications for cancer research. Just as it is inaccurate to assign a static function (e.g., apoptotic or anti-apoptotic) to a single protein, it is clear that static interpretations of mutations, that is, driver or passenger mutations, are also misleading. For example, given that the phenotypic role of JNK strongly depends on network state, it is clear that a mutation in JNK (and thus probably any other mutation) should not be statically labeled as a driver or passenger or as an oncogene or tumor suppressor, as such classifications are context dependent (e.g., disease or cell-type specific). Several examples, such as Myc[33] or WT1 (ref. 34) gene products that act as both tumor suppressors and oncogenes, support this idea. These results underscore the importance of assessing mutations based on their effects on signaling networks and of developing novel classification methods to do so. Along these lines, MAP2K4 (one of the protein kinases that can phosphorylate and activate JNK) has been shown to be recurrently lost or mutated in several cancers[35–38]. These represent prime examples of mutations that may display ambivalent phenotypic impact similar to JNK.

Motivated by the example of MAP2K4 and many other mutated kinases[38], we maintain that mutations capable of affecting signaling networks—which we call network-attacking mutations (**Fig. 2c**)—are more likely to affect phenotype than other mutations. Thus, we discuss a general strategy in which mutations in individual cancers are assessed based on, first, the likelihood they will affect protein function, and second, the cellular role of the signaling network that they are operating within (**Fig. 3**). Our strategy extends the concepts introduced by Waddington and elaborated by Kauffman and Huang *et al.*[31,32,39,40], where cancer mutations are turned into perturbations capable of reshaping these landscapes. We represent the cellular response or phenotype as another dimension where each network state (every point in the landscape) is constantly projected to and translated into a cellular decision or phenotypic outcome.

We postulate that network-attacking mutations affect the cell not by perturbing how the signaling landscape is projected to the phenotypic dimension, but by changing the ensemble of dynamic networks that can be manifested in a cell and, in consequence, the number and stability of steady states in the signaling landscape, thus creating new attractor states that only cancer cells can occupy, also known as cancer network attractors (**Fig. 3**). This has additional implications for other mechanisms, such as oncogene and non-oncogene addiction[41], where cancer cells would be trapped in cancer attractor states and could escape from them by reverting the genomic aberration that initially

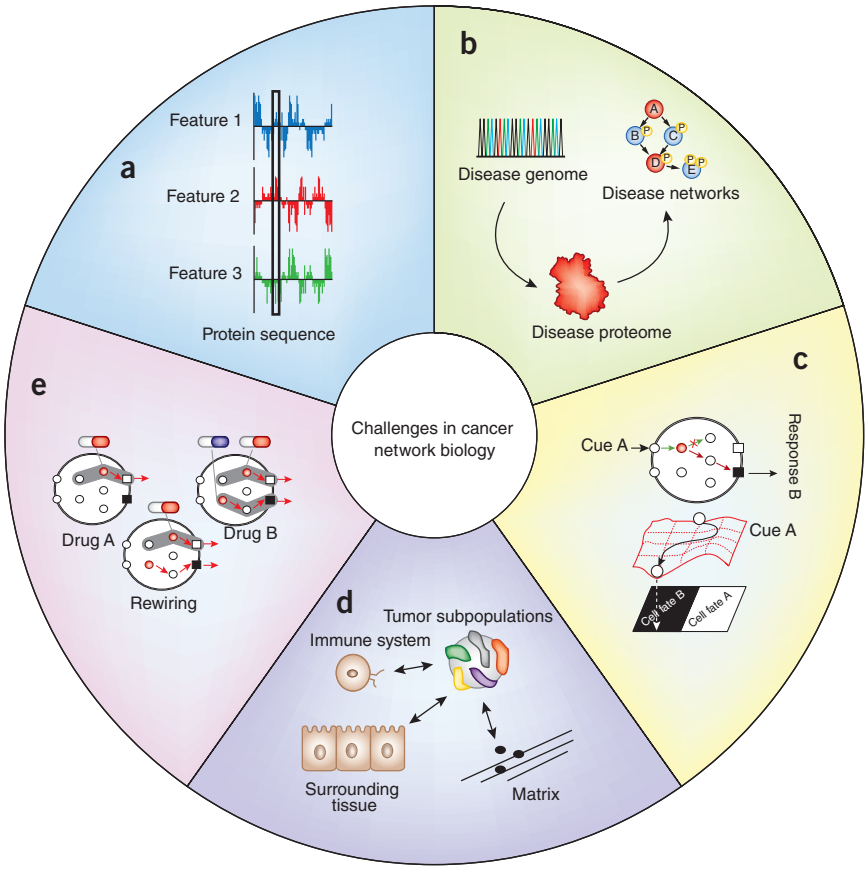


Figure 2 Challenges in cancer network biology. (**a**) Functional consequences of cancer mutations. Using an ensemble of protein-function features (e.g., ATP binding, substrate specificity, activation of the protein kinase or phospho-tyrosine binding), which together represent a comprehensive description of a protein's molecular functions, will enable more accurate and predictive evaluation of cancer mutations. (**b**) Modeling of disease networks. Although experimental and computational tools for modeling molecular networks exist, creating more comprehensive, sensitive and accurate new tools especially designed to model disease-associated networks still represents a big challenge in network biology. (**c**) Network-attacking mutations and cancer network attractors. Network-attacking mutations are mutations that lead to a new cellular phenotype by perturbing signaling networks either at the network structure or the network dynamics level. Network-attacking mutations transform signaling networks, generating new possible network states by changing the number and/or stability of steady states in the signaling landscape[31,32,39,40]. These acquired signaling capabilities lead to alterations in the cell's normal 'cue-signal-output' flow and thereby drive disease phenotypes (see **Fig. 3** for further details). (**d**) Tumor subpopulations and micro-environment. The field is only beginning to comprehend the complex interactions that exist between different co-evolving tumor cell subpopulations and between those cells and the tumor microenvironment, both of which strongly influence tumor progression. (**e**) Network-aware and temporal drugs. As predicted by R.L. and Pawson[66] several years ago, new pharmaceutical strategies that target networks instead of single proteins are becoming available[47,48]. We predict this trend will not only continue, but also include recent advances that highlight the possibility to 'cure' networks using time- and order-dependent therapies[68]. In coming years, the discovery of resistant, metastatic, tissue or cell-specific networks could lead to an even greater advance in the field of network medicine (**Fig. 5**).

caused the perturbed landscape. Given the high degree of determinism that exists between signaling networks, landscapes and phenotypes, we argue that network-attacking mutations are at the heart of all new decision-making capabilities acquired by cancer cells. Consequently, in our view, the study of both network-attacking mutations and new attractor states acquired by cancer cells, that is, cancer network attractors, deserves the highest priority from the field. Such studies should be performed through systematic and quantitative sampling of cell dynamics at multiple levels (e.g., genomic or epigenetic, proteomic and phenotypic), followed by nonlinear interpolation and integrative computational modeling (**Fig. 4**).

The first network-attacking cancer mutation, described more than 15 years ago[42], was a point mutation in the kinase domain of *RET* (M918T), which leads to a switch in peptide specificity. In line with their importance, network-attacking mutations have attracted more attention in recent years[43–48]. Moreover, information has been accumulating steadily about how specificity in signaling networks and modular protein domains emerges[49–51], leading to the definition of determinants of specificity in protein domains[52,53]. These determinants, sometimes referred to as specificity-determining residues, are residues that can lead to substrate specificity changes after mutation. Notably, direct mutagenesis of these determinants of specificity has been used to rewire the entire histidine kinase signaling system in bacteria in a predictive manner[54]. Recent follow-up work indicates that mutations in determinants of specificity prevent cross-talk and allow protein family expansions[55], in a process similar to the one powered by negative selection over Src homology 3 (SH3) protein domains that show similar specificity[56]. We propose that similar studies in human signaling networks, coupled with mapping of cancer mutations on these determinants of specificity, would shed new light on whether signaling rewiring is a general principle of oncogenesis and tumor progression, knowledge of which would in turn be critical as molecular therapies target proteins and their networks and not genes.

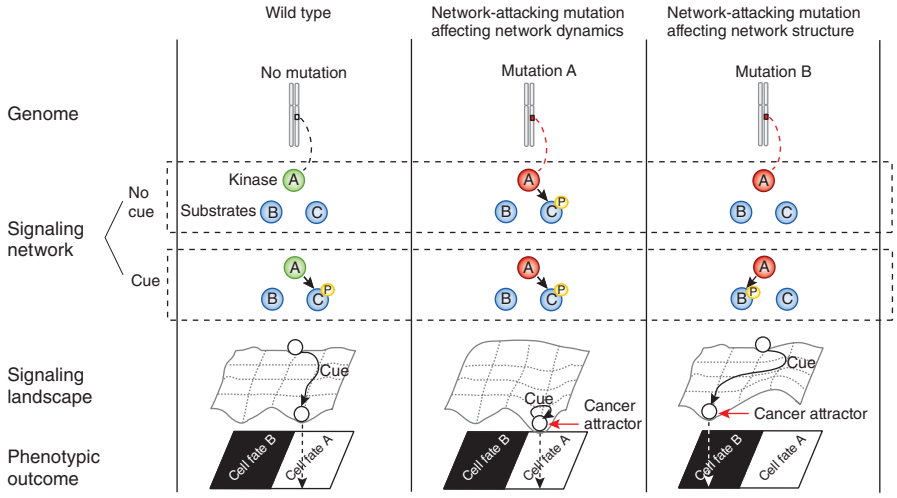


Figure 3 Network-attacking cancer mutations. Proteins are the key elements of signaling networks as a result of their ability to integrate external cues and direct the information flow toward a specific cellular outcome (e.g., epidermal growth factor (EGF) leading to proliferation or tumor necrosis factor alpha (TNF-α) leading to apoptosis). Network-attacking mutations affect the 'cue-signal-output' cellular information flow by affecting either the dynamics (middle), for example, by keeping proteins constitutively active, or the structure (right), by affecting protein specificity, of the signaling networks. Signaling networks can be represented as a landscape with the most likely network states represented as valleys (stable steady states or attractors) and the least likely network states as mountains (unstable steady states). Network-attacking mutations dysregulate signaling networks by perturbing the number and/or stability of steady states in the landscape, effectively creating new cancer-specific attractors that only cancer cells will be able to reach.

Despite the fact that the number of known cancer network-attacking mutations is still relatively low, recent findings suggest that in-frame mutations are enriched on interaction interfaces[57], which implies they are also likely to affect determinants of specificity. Moreover, many fusion proteins have been discovered that likely directly rewire or create new network states[58]. Given the rate at which cancer mutations are being reported and the development of new computational methods for systematically identifying these mutations (**Fig. 2b**), we predict a steep increase in the number of network-attacking mutations that will be uncovered in the coming years.

Personalized cancer network biology

Led by recent advances in sequencing technologies, the amount of data on cancer genome mutations is growing exponentially[59]. Current efforts

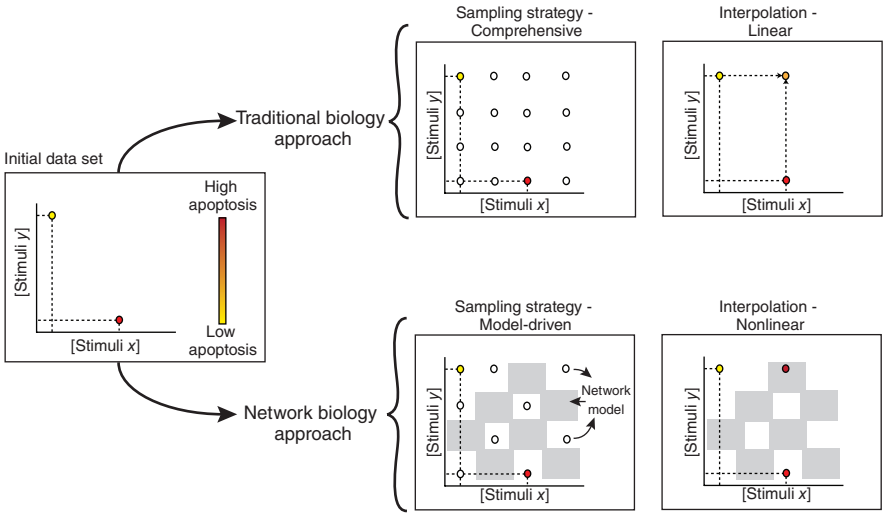


Figure 4 Traditional versus network biology approaches. In more traditional biological approaches, where only one or a few genes or proteins are sampled across a limited set of conditions, there has been limited success in deriving predictive models across conditions or cell types that would require comprehensive sampling. In contrast, network biology relies on systematic sampling across combinations of states that result in increased performance of a network model. Unlike classic approaches, in which the system is stimulated with single specific cellular cues (e.g., growth factor), in the network biology approach, the multivariate nature of signaling networks and the nonlinear relationship between signaling input and output can be successfully elucidated by interrogating the system with multiple orthogonal cues.

from the Cancer Genome Atlas and Cancer Genome Project, now under the umbrella of the International Cancer Genome Consortium[60], will facilitate the annotation and collection of cancer genome data. We foresee similar waves of technological progress and the generation of new consortiums in the cancer proteomics fields in the near future. The establishment of the Clinical Proteomic Tumor Analysis Consortium (http://proteomics.cancer.gov/programs/cptacnetwork), and the implementation of new approaches[61] and labeling techniques[62] optimized for patient samples are encouraging advances in this direction.

These advances, however, will need to be coordinated with new algorithmic and experimental high-throughput methods (e.g., high-content screening) capable of interpreting this flood of information because the functional interpretation of the data is currently the main bottleneck in the field of personalized cancer network biology. Computational integration of large quantitative data sets is also becoming increasingly important, and thus there is a growing requirement for supercomputing infrastructure with large algorithmic dynamic range (e.g., next-generation large shared memory systems). Benchmarking and validation of systematic workflows and algorithms is already receiving increasing attention through initiatives, such as the DREAM challenge[63] and IMPROVER[64].

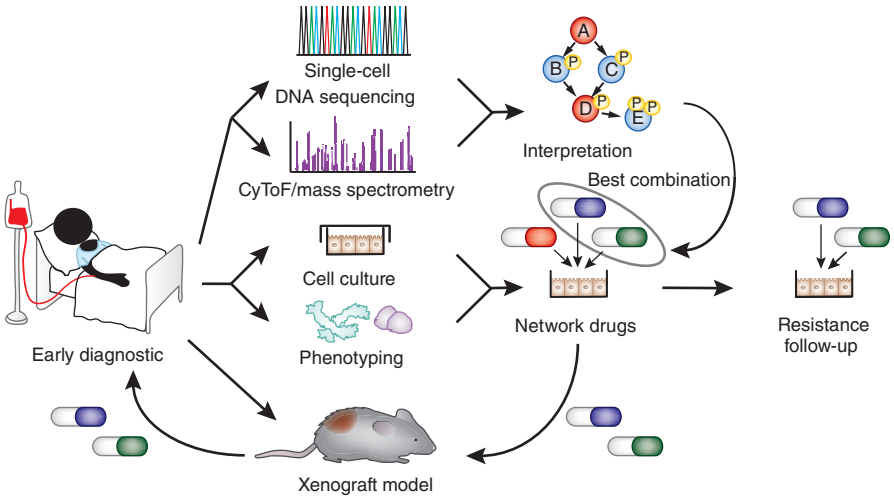


Figure 5 Personalized cancer network biology. The goal of personalized cancer network biology is to be able to treat each tumor with the best combination of drugs tailored to that tumor. Ideally, early diagnosis should be followed by the development of tumor-specific cell lines and xenograft models, cancer genome sequencing, and proteomic and phenotypic analysis. Combinations of network drugs should then be tried in the tumor-specific cell line and xenograft model and eventually transferred back to the patient. Continuing to treat the tumor-specific cell culture with the same network drug combination as is used in the patient may be useful for understanding potential resistance and/or metastasis.

Two emerging areas in network biology that are likely to contribute to the future of cancer research are the study of cell-cell interactions (**Fig. 2d**) and drugs specifically designed to interfere with diseased network dynamics (that is, network drugs; **Fig. 2e**).

R.L. and collaborators[65] studied cell-cell interactions by isotopically labeling two distinct subpopulations of cells, one expressing ephrin-B1$^+$ and the other Eph-B2$^+$, and carrying out a comprehensive phosphoproteomic analysis. This strategy facilitated the first measurements of phosphorylation events during the interaction of two cell subpopulations. The proliferative behavior of cancer cells is still poorly understood in part because it is difficult to experimentally study the transmission of proliferative factors from one cell to its neighbors[3]. Therefore, we argue that a similar isotopic labeling strategy could be used to investigate the cooperation between cells with different oncogenic lesions that together (that is, synthetic oncogenes; **Figs. 1b** and **2d**) lead to tumor formation[11].

Combination drugs that interfere with disease networks (so-called network medicine[66]) have been shown to lead to a better response than single-hit therapies by causing secondary perturbations to signaling networks[47,48,67]. Recent work by the Yaffe laboratory represents a clear leap forward within the field of network medicine[68,69]. Following network modeling, Yaffe and colleagues[68] managed to decode the signaling network dynamics that drive resistance to DNA-damaging chemotherapy. This information was used to sensitize otherwise resistant triple-negative breast cancer cells to conventional DNA-damaging chemotherapy by administering doxorubicin (Adriamycin, Doxil) and erlotinib (Tarceva) in an order- and time-dependent fashion. This could be considered the first example of temporal network drugs (**Figs. 2e** and **5**).

We predict that personalized or even tumor-specific cancer therapy will become a reality in the foreseeable future, starting from early diagnosis of the disease, followed by next-generation sequencing, proteomic analysis, high-throughput profiling of phenotypic cell states in the tumor and design of patient-specific combinations of network drugs with resistance follow-up (**Fig. 5**). Relatively new techniques, such as single-cell and high-depth sequencing[70,71], imaging[72] and cytometry time-of-flight[73], could prove especially valuable for monitoring the number, properties and behavior of different tumor subclones (**Fig. 2d**). Ideally, network drugs, such as the aforementioned order- and time-dependent combination[68], should then be chosen based on the interpretation of sequencing as well as the proteomic and phenotypic analysis of tumor cells and tested on the tumor-specific cell lines and xenograft model. The best-performing combination should ultimately be transferred back to the patient (**Fig. 5**). This whole process should take the shortest time possible to avoid the evolution of the tumor in the patient and the consequent loss of relationship between the primary tumor and the cell line. Tumor-specific cell lines would be kept and treated with the same drugs used in the patient to monitor tumor evolution and treat for resistance and/or metastasis as soon as there is enough evidence of it (**Fig. 5**). Ideally, every patient and paired xenograft or cell line should have a complete electronic record showing the treatment history to facilitate retrospective and cross-disease studies[74,75].

Conclusions

Although we have highlighted some of the challenges that still exist in cancer network biology, substantial progress is also being made. For example, the usage of patient-derived tumor tissue in animal xenograft models to test the response to particular drugs aimed at developing new personalized cancer therapy is rapidly becoming an established technology[76]. Surgical orthotopic implantation to transplant tumors taken directly from the patient to the corresponding organ of immunodeficient mice[77] is currently one of the most promising methods to enable drug screening in patients. In addition, new clinical trials, such as the MD Anderson T9 project[78], are under way in which patients are given therapy that targets tumor-specific aberrations. Nevertheless, the implementation of the strategy depicted in **Figure 5** would benefit from further developments in technology, funding and legislation. For

example, generating models for cancer research that represent human patient diversity[79] and mimicking the complexity of tumor microenvironments (J.T.E. and collaborators)[80] remain extraordinary challenges (**Fig. 2**), and further research efforts and investments are required. As cancer biology becomes a 'big data' science, similar to physics, we expect to see more systematic, data-driven research efforts that will uncover and confront many of the tumor complexities that have remained elusive so far.

Despite recent predictions of >13 million cancer deaths in 2030 (ref. 81), as discussed in this Perspective, we foresee that within this timeframe tumor-specific medicine will become a reality, thanks to a new generation of cancer network biologists who will hopefully overcome these challenges, positively contributing to the battle against this devastating disease and the significant reduction of patient suffering.

ACKNOWLEDGMENTS

We apologize to our colleagues whose work could not be cited due to space limitations. We thank all members of the C-SIG (DTU), the ErlerLab (BRIC), M. Yaffe (MIT) and N. Brunner (KU) for critical input on this manuscript. R.L. is a Lundbeck Foundation Fellow and is supported by a Sapere Aude Starting Grant from The Danish Council for Independent Research and a Career Development Award from Human Frontier Science Program. J.T.E. is supported by a Hallas Møller Stipend from the Novo Nordisk Foundation. Visit http://www.networkbio.org/, http://www.lindinglab.org/ and http://www.erlerlab.org/ for more information on cancer-related network biology.

AUTHOR CONTRIBUTIONS

All correspondence should be addressed to both J.T.E. and R.L.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nbt.2345.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Nash, J.F. Equilibrium points in N-person games. *Proc. Natl. Acad. Sci. USA* **36**, 48–49 (1950).
2. Nash, J.F. Non-cooperative games. *Ann. Math.* **54**, 286–295 (1951).
3. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
4. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
5. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
6. Stratton, M.R., Campbell, P.J. & Futreal, P.A. The cancer genome. *Nature* **458**, 719–724 (2009).
7. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
8. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
9. Hou, Y. *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **148**, 873–885 (2012).
10. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
11. Wu, M., Pastor-Pareja, J.C. & Xu, T. Interaction between RasV12 and scribbled clones induces tumour growth and invasion. *Nature* **463**, 545–548 (2010).
12. Ng, P.C. & Henikoff, S. Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.* **7**, 61–80 (2006).
13. Dixit, A. *et al.* Sequence and structure signatures of cancer mutation hotspots in protein kinases. *PLoS ONE* **4**, e7485 (2009).
14. Pazos, F. & Bang, J.-W. Computational prediction of functionally important regions in proteins. *Curr. Bioinform.* **1**, 15–23 (2006).
15. Fowler, D.M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **7**, 741–746 (2010).
16. Jensen, L.J. *et al. Ab initio* prediction of human orphan protein function from post-translational modifications and localization features. *J. Mol. Biol.* **319**, 1257–1265 (2002).
17. Socolich, M. *et al.* Evolutionary information for specifying a protein fold. *Nature* **437**, 512–518 (2005).
18. Russ, W., Lowery, D., Mishra, P., Yaffe, M. & Ranganathan, R. Natural-like function in artificial WW domains. *Nature* **437**, 579–583 (2005).
19. Puntervoll, P. *et al.* ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* **31**, 3625–3630 (2003).
20. Lim, W.A. & Pawson, T. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* **142**, 661–667 (2010).
21. Seet, B.T., Dikic, I., Zhou, M.M. & Pawson, T. Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* **7**, 473–483 (2006).
22. Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein sectors: Evolutionary units of three-dimensional structure. *Cell* **138**, 774–786 (2009).
23. Reynolds, K.A., McLaughlin, R. & Ranganathan, R. Hot spots for allosteric regulation on protein surfaces. *Cell* **147**, 1564–1575 (2011).
24. Wan, P.T. *et al.* Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**, 855–867 (2004).
25. Janes, K.A. *et al.* A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* **310**, 1646–1653 (2005).
26. Lei, K. & Davis, R.J. JNK phosphorylation of Bim-related members of the Bcl2 family induces Bax-dependent apoptosis. *Proc. Natl. Acad. Sci. USA* **100**, 2432–2437 (2003).
27. Lamb, J.A. *et al.* JunD mediates survival signaling by the JNK signal transduction pathway. *Mol. Cell* **11**, 1479–1489 (2003).
28. Abreu-Martin, M.T. *et al.* Fas activates the JNK pathway in human colonic epithelial cells: lack of a direct role in apoptosis. *Am. J. Physiol.* **276**, G599 (1999).
29. Jeong, H., Mason, S.P., Barabasi, A.L. & Oltvai, Z.N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
30. Shah, S.P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* advance online publication, doi:10.1038/nature10933 (4 April 2012).
31. Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **128**, 11–45 (1987).
32. Kauffman, S.A. & Weinberger, E.D. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.* **141**, 211–245 (1989).
33. Uribesalgo, I., Benitah, S.A. & Di Croce, L. From oncogene to tumor suppressor: The dual role of Myc in leukemia. *Cell Cycle* **11**, 1757–1764 (2012).
34. Yang, L., Han, Y., Saurez Saiz, F. & Minden, M.D. A tumor suppressor and oncogene: the WT1 story. *Leukemia* **21**, 868–876 (2007).
35. Ellis, M.J. *et al.* Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* **486**, 353–360 (2012).
36. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
37. Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466**, 869–873 (2010).
38. Greenman, C. *et al.* Pattern of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
39. Waddington, C.H. *The Strategy of the Genes: a Discussion of Some Aspects of Theoretical Biology* (Allen & Unwin, 1957).
40. Huang, S. & Ingber, D.E. Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Exp. Cell Res.* **261**, 91–103 (2000).
41. Luo, J., Solimini, N.L. & Elledge, S.J. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* **136**, 823–837 (2009).
42. Songyang, Z. *et al.* Catalytic specificity of protein-tyrosine kinases is critical for selective signaling. *Nature* **373**, 536–539 (1995).
43. Zhong, Q. *et al.* Edgetic perturbation models of human inherited disorders. *Mol. Syst. Biol.* **5**, 321 (2009).
44. Dreze, M. *et al.* 'Edgetic' perturbation of a *C. elegans* BCL2 ortholog. *Nat. Methods* **6**, 843–849 (2009).
45. Pe'er, D. & Hacohen, N. Principles and strategies for developing network models in cancer. *Cell* **144**, 864–873 (2011).
46. Vidal, M., Cusick, M.E. & Barabási, A.-L.L. Interactome networks and human disease. *Cell* **144**, 986–998 (2011).
47. Schoeberl, B. *et al.* Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. *Sci. Signal.* **2**, ra31 (2009).
48. Huang, P.H. *et al.* Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc. Natl. Acad. Sci. USA* **104**, 12867–12872 (2007).
49. Miller, M.L.L. *et al.* Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.* **1**, ra2+ (2008).
50. Linding, R. *et al.* Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415–1426 (2007).
51. Mok, J. *et al.* Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. *Sci. Signal.* **3**, ra12 (2010).
52. Brinkworth, R.I., Breinl, R.A. & Kobe, B. Structural basis and prediction of substrate specificity in protein serine/threonine kinases. *Proc. Natl. Acad. Sci. USA* **100**, 74–79 (2003).
53. Turk, B.E. Understanding and exploiting substrate recognition by protein kinases. *Curr. Opin. Chem. Biol.* **12**, 4–10 (2008).
54. Skerker, J.M. *et al.* Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043–1054 (2008).
55. Capra, E.J., Perchuk, B.S., Skerker, J.M. & Laub, M.T. Adaptive mutations that prevent crosstalk enable the expansion of paralogous signaling protein families. *Cell* **150**, 222–232 (2012).
56. Zarrinpar, A., Park, S.H. & Lim, W.A. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **426**, 676–680 (2003).
57. Wang, X. *et al.* Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat. Biotechnol.* **30**, 159–164 (2012).
58. Brehme, M. *et al.* Charting the molecular network of the drug target Bcr-Abl. *Proc. Natl. Acad. Sci. USA* **106**, 7414–7419 (2009).
59. Wong, K.M.M., Hudson, T.J. & McPherson, J.D. Unraveling the genetics of cancer: genome sequencing and beyond. *Annu. Rev. Genomics Hum. Genet.* **12**, 407–430 (2011).

60. Ledford, H. Big science: the cancer genome challenge. *Nature* **464**, 972–974 (2010).
61. Bensimon, A., Heck, A.J.R. & Aebersold, R. Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* **81**, 379–405 (2012).
62. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J.R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Methods* **7**, 383–385 (2010).
63. Prill, R.J. *et al.* Towards a rigorous assessment of systems biology models: the DREAM3 challenges. *PLoS ONE* **5**, e9202 (2010).
64. Meyer, P. *et al.* Verification of systems biology research in the age of collaborative competition. *Nat. Biotechnol.* **29**, 811–815 (2011).
65. Jørgensen, C. *et al.* Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells. *Science* **326**, 1502–1509 (2009).
66. Pawson, T. & Linding, R. Network medicine. *FEBS Lett.* **582**, 1266–1270 (2008).
67. Chandarlapaty, S. *et al.* AKT inhibition relieves feedback suppression of receptor tyrosine kinase expression and activity. *Cancer Cell* **19**, 58–71 (2011).
68. Lee, M.J. *et al.* Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780–794 (2012).
69. Erler, J.T. & Linding, R. Network medicine strikes a blow against breast cancer. *Cell* **149**, 731–733 (2012).
70. Navin, N. *et al.* Tumor evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
71. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
72. Pedersen, M.W. *et al.* Sym004: a novel synergistic anti-epidermal growth factor receptor antibody mixture with superior anticancer efficacy. *Cancer Res.* **70**, 588–597 (2010).
73. Bendall, S.C. & Nolan, G.P. From single cells to deep phenotypes in cancer. *Nat. Biotechnol.* **30**, 639–647 (2012).
74. Roque, F.S. *et al.* Using electronic patient records to discover disease correlations and stratify patient cohorts. *PLOS Comput. Biol.* **7**, e1002141 (2011).
75. Jensen, P.B., Jensen, L.J. & Brunak, S. Mining electronic health records: towards better research applications and clinical care. *Nat. Rev. Genet.* **13**, 395–405 (2012).
76. Blumenthal, R.D. & Goldenberg, D.M. Methods and goals for the use of in vitro and in vivo chemosensitivity testing. *Mol. Biotechnol.* **35**, 185–197 (2007).
77. Hoffman, R.M. Orthotopic mouse models expressing fluorescent proteins for cancer drug discovery. *Expert Opin. Drug Discov.* **5**, 851–866 (2010).
78. Gonzalez-Angulo, A.M., Hennessy, B.T. & Mills, G.B. Future of personalized medicine in oncology: a systems biology approach. *J. Clin. Oncol.* **28**, 2777–2783 (2010).
79. Hunter, K.W. Mouse models of cancer: does the strain matter? *Nat. Rev. Cancer* **12**, 144–149 (2012).
80. Cox, T.R. & Erler, J.T. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. *Dis. Model. Mech.* **4**, 165–178 (2011).
81. WHO. World health organization fact sheet 297 (2012). http://www.who.int/mediacentre/factsheets/fs297/en/

Chapter II

Part II

Experimental and Computational Tool for Analysis of Signaling Networks in Primary Cells

Experimental and Computational Tools for Analysis of Signaling Networks in Primary Cells

Erwin M. Schoof[1] and Rune Linding[1]

[1]Cellular Signal Integration Group (C-SIG), Center for Biological Sequence Analysis (CBS), Department of Systems Biology, Technical University of Denmark (DTU), Lyngby, Denmark

Cellular information processing in signaling networks forms the basis of responses to environmental stimuli. At any given time, cells receive multiple simultaneous input cues, which are processed and integrated to determine cellular responses such as migration, proliferation, apoptosis, or differentiation. Protein phosphorylation events play a major role in this process and are often involved in fundamental biological and cellular processes such as protein-protein interactions, enzyme activity, and immune responses. Determining which kinases phosphorylate specific phospho sites poses a challenge; this information is critical when trying to elucidate key proteins involved in specific cellular responses. Here, methods to generate high-quality quantitative phosphorylation data from cell lysates originating from primary cells, and how to analyze the generated data to construct quantitative signaling network models, are presented. These models can subsequently be used to guide follow-up in vitro/in vivo validation studies. *Curr. Protoc. Immunol.* 104:11.11.1-11.11.23. © 2014 by John Wiley & Sons, Inc.

Keywords: phosphorylation • mass spectrometry • network biology • primary cell signaling

CELL SIGNALING IN PRIMARY CELLS

Cellular responses to environmental stimuli are driven primarily by information processing in signaling networks. Cells receive multiple input cues simultaneously at any given time, and have to decide on appropriate cellular responses such as apoptosis, proliferation, differentiation, or migration (Manning et al., 2002; Jorgensen et al., 2009). Post-translational modifications (PTMs) are an important mechanism for cells to accomplish this, as they alter protein activity and interactions as required by a given cellular response. For example, PTMs can direct and modulate the binding of protein domains to a specific motif on a substrate protein (Pawson, 1995; Seet et al., 2006), thereby altering the kinetics of a protein-protein interaction. Although many types of PTMs exist, such as ubiquitination, acetylation, or methylation, phosphorylation events are among the most extensively used by the cell. They largely govern cellular information processing, and have been demonstrated to be involved in most fundamental biological and cellular processes such as protein-protein interactions, enzyme activity, and immune response (Miller and Berg, 2002; Cannons and Schwartzberg, 2004; Seet et al., 2006; Readinger et al., 2009).

When considering phosphorylation-based signaling within the immune system, it is well established that many immune responses are evoked through the activation of specific receptors on the cell surface by, for example, ligand or antigen binding. These, in turn, activate protein kinases to generate a cellular response through specific phosphorylation network dynamics. For example, Syk, Tec, Src, and protein kinase C (PKC) family kinases have extensively demonstrated to be involved in immune responses involving T-cell activation upon antigen presentation (Monks et al., 1998; Isakov and Altman,

Current Protocols in Immunology 11.11.1-11.11.23, February 2014
Published online February 2014 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/0471142735.im1111s104

2002; Miller and Berg, 2002). Due to its integral role in overall cellular functioning, dysregulation of phosphorylation-based signaling often causes severe changes to the cellular phenotype by evoking distinct alterations to normal cellular responses. This, combined with their ubiquitous nature, implicates them in many human diseases, and the modulation of their dynamics constitutes potential treatment targets (Shawver et al., 2002; Tan et al., 2009; Fedorov et al., 2010; Lemmon and Schlessinger, 2010).

To establish causal relationships between observed phosphorylation events and their effects on signaling networks, one must decipher not only the kinase-phosphosite relationships (i.e., which kinase(s) phosphorylate(s) which phosphorylation sites/substrates), and which phosphatases and phospho-binding domains (e.g., SH2, BRCT, or PTB domains) dephosphorylate and interact, respectively, with the observed phosphorylation sites. Additionally, insight must be gained into the biochemical effects that the modification of these sites exerts on the cellular signaling proteins and networks and, ultimately, how these alter cellular phenotypes or behavior (Cantley et al., 1991; Pawson and Hunter, 1994; Pawson, 1995; Pawson and Kofler, 2009; Brognard and Hunter, 2011; Creixell et al., 2012).

Here, methods to generate high-quality, quantitative phosphorylation data from cell lysates originating from primary cells, such as monocyte-derived immature dendritic cells, are described. The strategy for accomplishing this involves: (1) performing cell lysis, protein digestion, and peptide labeling (see Basic Protocol 1); (2) separating the peptides according to charge state, and allowing the fractions to be subsequently enriched for phosphopeptides separately (SCX fractionation; see Basic Protocol 2); (3) performing specific enrichment techniques that need to be deployed in order to boost the detection of phosphopeptides (see Basic Protocol 3); (4) purifying samples for MS analysis (see Basic Protocol 4); and (5) analyzing the generated data to construct quantitative signaling network models, which can be used to guide follow-up in vitro/in vivo validation studies (see Basic Protocol 5).

Several hurdles must be overcome when studying phosphorylation-based signaling (in primary cells). First, the intrinsically low signal-to-noise ratio of phosphorylation events due to their low abundance and low stoichiometry compared to non-phosphorylated peptides (Jin et al., 2010) represents a challenge that significantly increases the complexity of detecting these events. A second challenge is the transient nature of kinase-substrate interactions, which, due to the high off-rate (k_{off}) of a kinase-substrate interaction, often renders it infeasible to determine experimentally the substrates of a particular kinase using conventional affinity-based biochemistry methods such as tandem affinity purification (TAP) or immunoprecipitation (IP) MS (Burckstummer et al., 2006; Dyson et al., 2011). These approaches depend on stable interactions between the target proteins and the antibody to separate the antibody-bound proteins from the cell lysate. In this manner, one may be able to enrich for kinases and proteins bound to them, but this does not directly translate to the kinase phosphorylating these proteins, as they may purely exist in a scaffolding complex to bring the kinase in the appropriate cellular context for targeting other substrates. Similarly, in vitro kinase reactions do not reflect the cellular context, and thus the specificity in such assays and kinase peptide arrays do not accurately reflect cellular specificity and often leads to large amounts of false positives (Obenauer et al., 2003; Hjerrild et al., 2004). Kinases and substrates typically interact in a transient manner. This makes cellular (or so-called in vivo) kinase-substrate interactions challenging or impossible to capture by experimental methods alone (Linding et al., 2007).

While mass spectrometry (MS) is now able to identify and quantify thousands of phosphorylated residues from a single sample (Bodenmiller and Aebersold, 2010; Mohammed and Heck, 2011; Monetti et al., 2011; Munoz and Heck, 2011), thereby providing a

robust solution to the first aforementioned challenge (i.e., low signal-to-noise ratio), this technique often cannot solve the aforementioned second challenge (i.e., identifying the responsible kinases for these sites). Moreover, so-called Shokat kinases, which rely on a modified ATP binding pocket within the kinase domain in an attempt to utilize labeled ATP for identifying direct kinase substrates, in addition to their limited kinome-coverage, cannot be readily deployed in primary cells, as the cells need to be stably transfected to obtain the required kinase domain mutations (Shah and Shokat, 2003). This has led to a large knowledge gap between the identification of phosphorylation sites and their regulating kinases, information that is critical when attempting to elucidate kinase-substrate networks. It has thus been demonstrated that a combination of computational and experimental approaches is required. Computational approaches have been developed to address this issue, which, in combination with experimental techniques, can be deployed to decrease the knowledge gap (Linding et al., 2007; Miller et al., 2008; Szklarczyk et al., 2011).

GENERATING QUANTITATIVE PHOSPHO-PROTEOMICS DATA USING MASS SPECTROMETRY

While immunoblotting using phospho-specific antibodies was originally one of the most commonly used techniques to investigate phosphorylation events, the low-throughput nature of this approach, combined with its confined character (phosphopeptide-specific antibodies are required, biased by preconceived notions about which phosphorylation sites/proteins are important), non-linear dynamic range, and inaccurate quantitation, meant global quantitative approaches were desired. In the last decade, MS has been increasingly deployed, as it is able to routinely identify and quantify thousands of proteins in a single analysis, and is much more systematically biased (driven by protein stoichiometry and technical design of the instrument), allowing such biases to at least partially be corrected for (Callister et al., 2006; Prakash et al., 2007). Due to the low signal-to-noise ratio of phosphorylated peptides compared to the non-phosphorylated peptides, specific enrichment techniques need to be deployed in order to boost the detection of phosphopeptides. Several techniques exist for this, ranging from IP-based techniques using broad-spectrum phospho-specific antibodies (e.g., against phospho-tyrosine peptides or peptides with a simple motif, e.g., S/TQ for ATM/ATR kinases) to metal affinity-based approaches such as immobilized metal affinity chromatography (IMAC) or titanium dioxide (TiO_2; Kawahara et al., 1990; Tani and Suzuki, 1994; Posewitz and Tempst, 1999; Pandey et al., 2000; Jiang and Zuo, 2001; Tanl et al., 2002; Larsen et al., 2005; Rikova et al., 2007). These methods enable selective enrichment of phosphorylated peptides from a peptide pool, thereby making them more readily detectable for the mass spectrometer. An initial drawback of these approaches was the requirement of a relatively large amount of starting material. This has subsequently been overcome by steadily increasing enrichment efficiency as a result of technological developments, which currently makes it possible to identify several thousands of phosphopeptides from a few hundred micrograms of starting material (Engholm-Keller et al., 2012; Zhou et al., 2013). This, in turn, facilitates the investigation of the phosphorylation dynamics in biological systems where a limited number of cells are available, such as primary cells, cancer stem cells, or blood-circulating cells. Furthermore, sample fractionation techniques such as strong cation exchange (SCX) (Mohammed and Heck, 2011), hydrophilic interaction chromatography (HILIC) (McNulty and Annan, 2008), or electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) (Alpert, 2008) can spread the sample complexity across sample fractions, thereby facilitating greater phosphoproteome coverage by increasing the time available for the mass spectrometer to find unique peptides.

Due to the highly dynamic nature of biological systems, phosphorylation-based signaling networks, phosphoproteomes, or proteomes should not be conceptualized, interpreted,

nor described as static entities. Gaining a deeper understanding of the dynamics within signaling networks and how it relates to cell phenotypes is one of the major current challenges in systems biology (Creixell et al., 2012). To this end, it is important to elucidate the cellular information flow-through ensembles of signaling network states, which can be accomplished by conducting, e.g., time-series experiments or dose-response studies. The number and scale of time-points will depend on the system and biological question at hand, but dynamic monitoring of the system will generally give much more in-depth biological insight into the cellular processes driving a given phenotype (Janes et al., 2005; Miller-Jensen et al., 2007; Kreeger et al., 2010). This also enables one to explore the multivariate nature of cellular signaling (Linding, 2010; Jensen and Janes, 2012), which is based on the notion that cells have to integrate many signaling cues simultaneously, the responses to which are often non-linearly related to each other. This enables cells to integrate the different stimuli and respond with appropriate quantitative phenotypic outcomes. One can, for example, stimulate a biological system with a combination of stimuli, i.e., chemical inhibitors, RNAi, antigens, or antibodies (Pedersen et al., 2010), simultaneously or in a time-staggered manner for more comprehensive signaling network models to be constructed (Saez-Rodriguez et al., 2009). These can subsequently guide efforts to formulate so-called network-drugs, which target specific signaling network states rather than individual proteins (Pawson and Linding, 2008; Erler and Linding, 2010; Creixell et al., 2012; Lee et al., 2012).

In cell culture, a quantitative tool can easily be introduced through isotopic labeling, commonly known as stable isotope labeling by amino acids in cell culture (SILAC) (Ong et al., 2002). The principle in SILAC is the incorporation of non-radioactive isotopes through several (typically four to seven) cell divisions to ensure full isotope incorporation. While this is a very powerful approach for several cell types, it is not a suitable option for primary cells, as they can only undergo a limited, pre-determined number of divisions in culture, if any at all. Rather, a post-culture labeling method where proteins/peptides are labeled after cell lysis is a more effective approach. Several techniques for this exist, the primary ones being isobaric tag for relative and absolute quantitation (iTRAQ), tandem mass tag (TMT), or stable isotope dimethyl labeling (Thompson et al., 2003; Ross et al., 2004; Boersema et al., 2009). These labeling strategies all work based on the principle of adding a small but detectable mass shift to the cellular peptides to be able to mix, process, and subsequently analyze them together. The mass shift introduced by the labeling can be detected by the mass spectrometer, and used to trace the sample origin of a given peptide (e.g., to a specific time point, treatment condition, cell type, etc.). This also enables direct comparison of the abundance of the differentially labeled peptides, thus strengthening the quantitation of peptides. For simplicity, this unit focuses on the dimethyl-labeling method. The only limitation of this method compared to iTRAQ or TMT is that while the latter can be used to simultaneously label and compare up to eight samples simultaneously, dimethyl labeling is limited to triplex analysis. This does suffice for many experimental setups, however, and is comparable to the widely used SILAC approach. In general, dimethyl labeling is recommended as an appropriate and powerful default experimental approach for the study of signaling in primary cells, while more complex analyses (comparing more than three samples) would benefit from iTRAQ or TMT approaches.

Finally, this protocol focuses mainly on TiO_2-based enrichment, which, given the relatively higher abundance of phosphorylated serine (pSer) and threonine (pThr) residues in comparison to tyrosine phosphorylation (pTyr), will produce a larger number of pSer/pThr identifications than pTyr (Olsen et al., 2006). TiO_2-based enrichment gives rise to a significant global phosphoproteome coverage, while a fraction of the pTyr events can still be captured using TiO_2. However, pTyr enrichment using pTyr-specific antibodies such as pTyr-100/1000 or 4G10 is highly recommended if one desires a specific

focus on tyrosine kinases or pTyr signaling (Rikova et al., 2007; Jorgensen et al., 2009). In this case, the appropriate experimental protocols that are supplied by the antibody manufacturers are recommended. Nevertheless, despite the fact that TiO_2-based enrichment does not target tyrosine signaling specifically, it is still capable of identifying some of these events (Olsen et al., 2006). It is also worth pointing out that due to the high inter-connectedness of kinase-substrate signaling networks, pSer/pThr signaling events can still give insight to a particular phenotype in cases where high pTyr involvement is expected, as they are likely to also be utilized by the cell as "down-stream" effectors to achieve a specific response (Samelson et al., 1986; Dustin, 2009).

CELL LYSIS, PROTEIN DIGESTION, AND DIMETHYL LABELING

BASIC PROTOCOL 1

The following protocol can in principle be applied to any type of primary cells of interest, and should be done immediately after the experimental aim has been achieved (e.g., stimulation, mixing of cells, drug/antigen exposure) and preferably in a time-point-dependent manner. The number of cells to start with depends on the availability, but this protocol is optimized for protein amounts ranging from 2 to 24 mg of protein, or ~20 to 200 million cells. For an overview of the complete experimental workflow, see Figure 11.11.1; this protocol focuses on cell lysis, protein digestion, and peptide labeling.

Materials

Cell line(s) of interest
Phosphate-buffered saline (PBS; Sigma, cat. no. P5368), ice cold
Modified RIPA buffer (see recipe), ice cold
Acetone, HPLC-grade (Sigma, cat. no. 650501), −20°C
Denaturation buffer (see recipe)
Bradford reagent (Sigma, cat. no. B6916)
Dithiothreitol (DTT; Sigma, cat. no. 43815)
Chloroacetamide (CAA; Sigma, cat. no. 22790)
Lysyl endopeptidase (Lys-C; Wako, cat. no. 129-02541; 0.5 μg/μl stock solution made up in MilliQ water)
Triethyl ammonium bicarbonate (TEAB; Sigma, cat. no. T7408)
Trypsin (Sigma, cat. no. T6567; 0.5 μg/μl stock solution made up in 50 mM acetic acid)
Trifluoroacetic acid (TFA; Sigma, cat. no. T6508)
Acetic acid (Fisher Scientific, cat. no. A35-500)
Dimethyl labeling solution (see recipe)

15- or 50-ml tubes
Sonicator
Refrigerated centrifuge
Axial rotator
SepPak C18 columns (Waters, cat. no. WAT020515)
10-ml syringe (polypropylene)

Additional reagents and equipment for Bradford assay (Bradford, 1976)

Perform cell lysis and digestion

1. Remove the cell medium and wash cells two times with ice-cold PBS to remove any serum-containing medium. For adherent cells, pour out the medium, add ~20 ml of PBS for a 15-cm dish (use more or less according to culture vessel used), briefly swirl by hand, and discard. Repeat this process two times. For non-adherent cells, spin down cells 3 min at 300 × *g*, 5°C, in a 15-ml tube, remove the supernatant, and

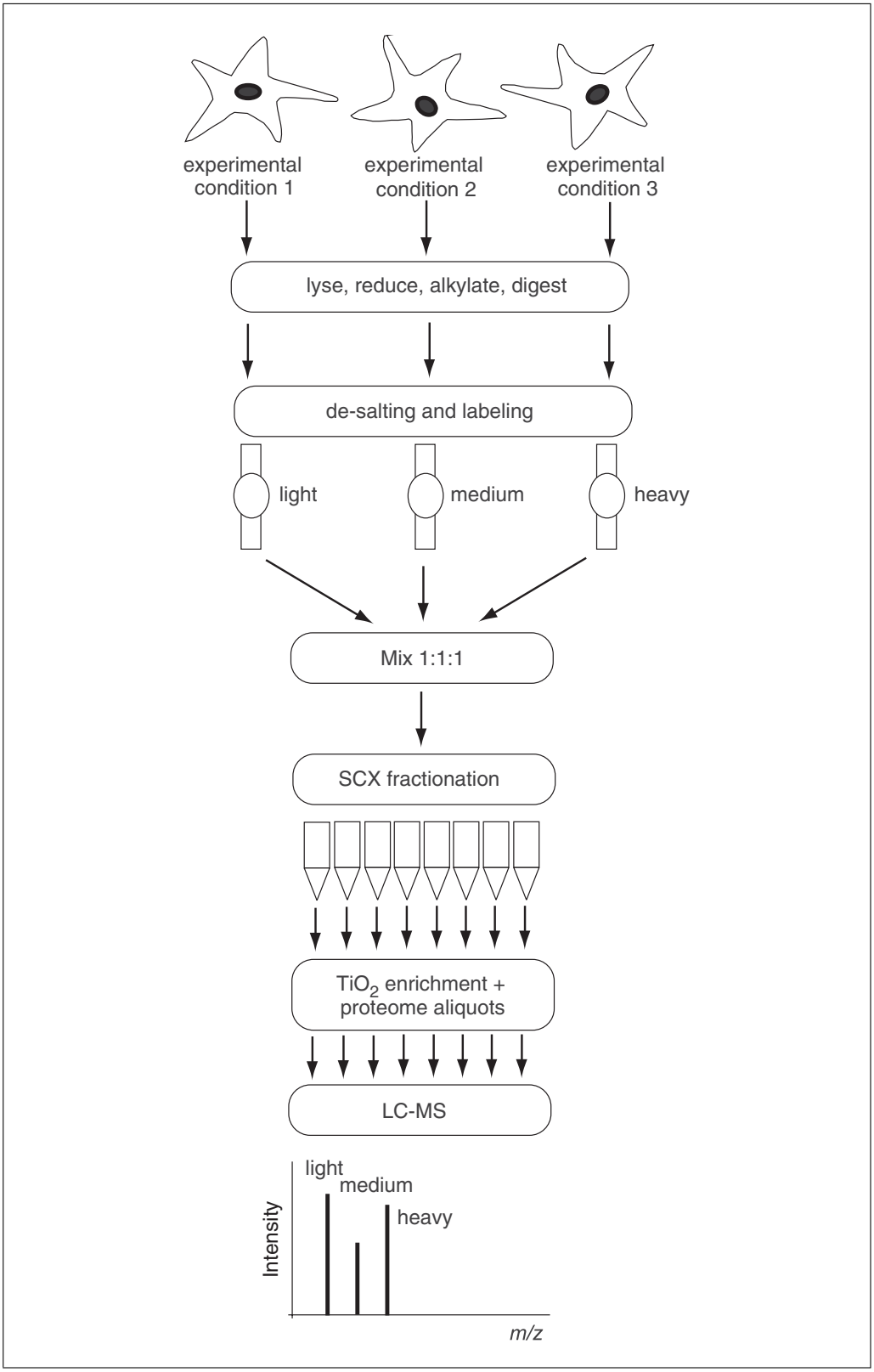


Figure 11.11.1 Experimental workflow overview, highlighting the key components of the sample preparation procedure.

add 10 ml of ice-cold PBS, pipetting up and down carefully. Repeat this process two times.

2. Remove the PBS from the final washing step, add 1 to 2 ml ice-cold RIPA buffer per 10×10^6 cells; if working with adherent cells, scrape the plates, otherwise pipet the cells and lysis buffer up and down until full lysis is achieved. Subsequently, transfer lysate to a 15- or 50-ml tube on ice (depending on total lysate volume), and sonicate on ice three times, 10 sec each time.
3. Centrifuge 20 min, full speed ~4500 × *g*, 4°C.
4. Transfer supernatant to a clean 50-ml tube, and add ice-cold acetone (−20°C) to a final concentration of ~80% acetone. Place at −20°C and precipitate proteins overnight.
5. Centrifuge 5 min at 2000 × *g*, 4°C, to pellet the proteins, and discard the acetone by decanting, being careful not to disturb the protein pellet.
6. Add sufficient denaturation buffer to a final concentration of ~5 to 10 mg/ml, and leave for a few hours to overnight at room temperature on an axial rotator to completely dissolve the protein pellet.

Determine protein concentration

7. Determine exact protein concentration using a Bradford assay (Bradford, 1976), either in cuvette- or 96-well plate format.
8. Add 1:1000 (v/v) of 1 M DTT to achieve a final concentration of 1 mM, and incubate 1 hr at room temperature on an axial rotator.
9. Add 1:100 (v/v) of 500 mM CAA to achieve a final concentration of 5 mM, and incubate 1 hr at room temperature in the dark on an axial rotator.
10. Check that the pH is 8, and add 1 μg of lysyl endopeptidase (Lys-C) per 100 μg of protein (1:100). For larger amounts of protein (>10 mg), add 1 μg of Lys-C per 200 μg of protein (1:200). Incubate ~4 to 5 hr at room temperature on an axial rotator.

    If the pH needs to be adjusted, use a very low volume of 1 M NaOH or HCl.

11. Dilute sample(s) 1:4 with 50 mM TEAB in water to reduce (Thio)urea concentration, and check that pH is 8.0 to 8.5 (adjust, if necessary, with 1 M NaOH or HCl).
12. Add 1 μg of Trypsin per 100 μg of protein (1:100). For larger amounts of protein (>10 mg), add 1 μg of Trypsin per 200 μg of protein (1:200). Incubate overnight at room temperature on an axial rotator.
13. Add TFA to a final concentration of 2% (using 20% TFA stock solution) to de-activate any remaining Trypsin, and centrifuge the acidified peptide mixture 5 min at 2000 × *g*, 20ºC, to clarify and transfer the supernatant to a clean tube.

De-salt samples and perform dimethyl labeling (adapted from Boersema et al., 2009)

If dimethyl labeling is not to be performed, skip step 19, the other steps must be performed for desalting purposes.

14. For each sample that is to be labeled, prepare a SepPak column by attaching a 10-ml syringe to it, after having removed the plunger.
15. Add 5 ml of 100% acetonitrile to each syringe, and allow it to run through the SepPak column by gravity.

If necessary and if no vacuum manifold is available, minimal pressure can be applied by replacing the plunger into the top of the syringe, but never push the plunger down beyond the rubber part sitting at the top of the syringe as this may put too much pressure on the column. The whole SepPak/dimethyl labeling process takes between 2 and 4 hr for optimal results.

16. Wash the SepPak column two times with 4 ml of 0.6% acetic acid solution each time, again allowing gravity to pull solution through the column.
17. Load equal amounts of each sample (as previously determined by the Bradford assay) onto their respective SepPak columns and allow gravity flowthrough; depending on the sample volume this can take a while.
18. Wash the SepPak column with 5 ml of 0.6% acetic acid solution.
19. Flush each SepPak column with 1 ml of its respective labeling reagent, repeating this procedure five times to ensure complete labeling.

    Again, this process could take a while and should take at least 10 min to ensure complete labeling.

20. Wash the SepPak column with 5 ml of 0.6% acetic acid solution.
21. Elute the labeled peptides from the SepPak column two times with 2 ml of 80% acetonitrile plus 0.6% acetic acid, each time.
22. Mix the differentially labeled samples, and proceed to Basic Protocol 2 for SCX fractionation, or Basic Protocol 3 if no SCX fractionation will be done (recommended for protein amounts <2 mg) and the sample will be directly enriched for phosphopeptides.

    In this case, if one is interested in analyzing proteome samples, other fractionation techniques such as gel-based fractionation (Schirle et al., 2003), HILIC fractionation (McNulty and Annan, 2008), or Offgel fractionation (Michel et al., 2003; Hörth et al., 2006) can be deployed to gain better proteome coverage.

BASIC PROTOCOL 2

SCX FRACTIONATION

To spread sample complexity over several fractions, protein samples >2 mg are recommended to be subjected to SCX fractionation. This will separate the peptides according to the charge state, and allow the fractions to be subsequently enriched for phosphopeptides separately, thereby gaining a better phosphoproteome coverage. This protocol covers sample injection, running the gradient, and subsequent pooling of fractions.

This protocol has been adapted from Olsen and Macek (2009).

Materials

Sample
Acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
SCX buffer A (see recipe)
SCX buffer B (see recipe)
Loading buffer: 1%TFA and 2% acetonitrile in MS H_2O

HPLC/FPLC system (e.g., GE Healthcare AktaMicro)
1-ml SCX column or equivalent (e.g., Resource S 1ml; GE Healthcare Resources)
2-ml microcentrifuge tubes

1. Load sample into the LC system as per manufacturer's instructions.
2. Load the peptides onto an equilibrated 1-ml SCX column as per manufacturer's instructions, and elute the peptides into clean 2-ml microcentrifuge tubes over a

30-min period using the following gradient: 1% to 30% SCX buffer B gradient, followed by 5 column volumes of 100% SCX solvent B, and ending the gradient with 5 column volumes of 100% SCX buffer A to equilibrate the column.

Make sure to collect all of the sample, including the flow-through and final equilibration fractions. Fractionation on the Resource S 1 ml column should be carried out at a flowrate of 1 ml/min.

3. Pool some of the fractions according to their chromatographic peaks, while obtaining about eleven pools of fractions, which can be individually enriched for phosphopeptides.

   The flow-through (early fractions) consists mainly of multiply phosphorylated peptides and will not bind to the SCX column; it is therefore recommended to pool, and sequentially enrich this pooled fraction for phosphopeptides at least three times.

4. If desired, dispense proteome samples into aliquots to be able to compare the phosphoproteome with the proteome.

   While exact amounts depend on the chromatography and amount of sample loaded, pipetting 5 to 10 μl from each pooled fraction is generally sufficient, and one should aim to have about six samples in total for MS analysis.

5. Acidify and reduce the acetonitrile concentration of the proteome samples with 100 μl of loading buffer, and keep for several hours at 4°C until the StageTipping stage; process as soon as possible.

TITANIUM DIOXIDE PHOSPHOPEPTIDE ENRICHMENT

BASIC PROTOCOL 3

This protocol covers the TiO_2-based enrichment procedure that enriches samples for phosphorylated peptides. These steps should be carried out at room temperature.

This protocol has been adapted from Thingholm et al. (2006) and Olsen and Macek (2009).

NOTE: This protocol is intended for SCX-fractionated samples. For non-fractionated samples, adjust step 3 to sequentially incubate the single sample three to five times separately to enrich for the majority of the phosphopeptides.

Materials

TiO_2 beads (GL Sciences, cat. no. 5020-75010)
TiO_2 loading solution (see recipe)
SCX samples (see Basic Protocol 2)
SCX buffer B (see recipe)
TiO_2 washing solution 1 (see recipe)
TiO_2 washing solution 2 (see recipe)
Acidification buffer (see recipe)
TiO_2 elution buffer 1 (see recipe)
TiO_2 elution buffer 2 (see recipe)

Automated sample shaker (e.g., Eppendorf Thermomixer)
End-over-end rotator
Centrifuge
C8 StageTips (Thermo Fisher, cat. no. SP321)
10-ml luer-lock syringes and StageTip adaptor (Millian, cat. no. HAM-31330)
96-well PCR plates
Vacuum centrifuge with microplate rotor (e.g., Thermo Savant SC250)
Litmus paper
Vortex

1. Make up the TiO_2-bead slurry solution by mixing ~1.5 mg TiO_2 beads per sample with 6 μl of TiO_2 loading solution, and put on an automated sample shaker (e.g., Eppendorf Thermomixer at 1400 rpm) for 15 min at room temperature. For example, when analyzing fifteen SCX fractions, mix 25 mg TiO_2 beads with 100 μl of TiO_2 loading solution.
2. Add 6 μl of the TiO_2 slurry to each sample, keeping the beads well suspended in the slurry in between sample loading, by briefly vortexing the slurry prior to transferring 6 μl to each sample. Incubate 30 min with end-over-end rotation at room temperature.
3. Centrifuge sample tubes 5 min at 2000 × *g*, room temperature, to pellet the TiO_2 beads, and for the most concentrated fractions (the flow-through and single-peak fractions, based on chromatography), transfer the supernatant to a clean tube and re-incubate with an additional 6 μl of TiO_2 slurry for 30 min. For all other fractions, aspirate off the supernatant, resuspend pellet in 100 μl of SCX buffer B, and transfer to a clean microcentrifuge tube. Keep at 4°C while the other samples are incubating, repeating this process until the flow-through has been enriched three to five times, each time storing the beads in a clean microcentrifuge tube for MS sample preparation.
4. Centrifuge all samples 5 min at 800 × *g*, room temperature, and aspirate supernatant.
5. Resuspend beads in 100 μl TiO_2 washing solution 1.
6. Centrifuge all samples 5 min at 800 × *g*, room temperature, and aspirate supernatant.
7. Resuspend samples in 50 μl TiO_2 washing solution 2, and transfer each sample to a separate C8 StageTip, pipetting sample onto the top of the pipet tip in order for the beads to collect on top of the C8 filter.
8. Flick the sample down into the StageTip using a wrist motion, and push the TiO_2 washing solution 2 through the filter using a syringe, leaving only the TiO_2 beads behind.
9. Pipet 40 μl of acidification buffer into one well for each sample of a 96-well PCR plate, as this improves phosphopeptide stability. Elute the phosphopeptides into the PCR plate (one well per C8 StageTip) using one application of 20 μl TiO_2 elution buffer 1, and one application of 20 μl TiO_2 elution buffer 2.
10. Vacuum centrifuge samples for ~55 min (time is dependent on the model of vacuum centrifuge used), without heat, until the total volume for each sample is ~20 μl. While waiting for this step to complete, one can prepare the C18 StageTips for final peptide purification before MS analysis according to Basic Protocol 4 (up to step 4).
11. Add 20 μl of acidification buffer, and check that pH <2 using litmus paper. In case of high pH (due to, e.g., insufficient ammonia removal during SpeedVac), add an additional 20 μl of acidification buffer until the pH is <2.
12. Cover the PCR plate, briefly vortex (not too vigorously), and centrifuge 1 min (without vacuum) to get the entire sample down into the well.

MASS SPECTROMETRY SAMPLE PREPARATION

BASIC PROTOCOL 4

Following completion of Basic Protocols 1 through 3, the samples are ready to be purified for MS analysis using C18 StageTips (Rappsilber et al., 2007).

Materials

Methanol, HPLC-grade (Sigma, cat. no. 34860)
Buffer B (see recipe)

Sample buffer (see recipe)
Samples (see Basic Protocol 1, 2, or 3)
Buffer A (see recipe)
Loading buffer (see recipe)

C18 StageTips (Thermo Fisher, cat. no. SP301)
10-ml luer-lock syringes and StageTip adaptors (Millian, cat. no. HAM-31330)
Vacuum centrifuge (e.g., Thermo Savant SC250)
Mass spectrometer with nanospray source (e.g., Thermo Fisher Q Exactive or Orbitrap Fusion)

1. Clearly label each C18 StageTip for the sample that is to be loaded onto it.
2. Prime the StageTips with 20 µl methanol, flicking the StageTip down using a wrist motion to get the liquid down into the C18 filter, and subsequently slowly pushing through the liquid using a syringe. Always ensure that a small amount of liquid remains on top of the filter to keep it from drying out.

   If many samples are to be prepared, one can opt to use a microcentrifuge for spinning the liquid through the C18 filter. In this case, place the StageTip into a pipet adaptor placed inside an empty 2-ml microcentrifuge tube. Spin the tips at ~800–1000 × g to allow the liquid to spin through in ~30 sec.

3. Push 20 µl of buffer B through the StageTips.
4. Wash StageTips two times with 20 µl sample buffer, each time.
5. Slowly push the previously prepared (phospho-) peptide samples through the StageTips.
6. Wash the StageTips two times with 20 µl buffer A, each time.

   At this stage, the samples can be stored at 4°C, as long as the C18 filter remains covered in buffer A. For phosphopeptide samples, the StageTips should not be stored for longer than 1 to 2 weeks, whereas proteome samples can be stored for weeks. Long-term storage (several months) of both types of samples can be done at −80°C.

7. Just before MS analysis, elute the purified StageTips two times with 20 µl buffer B, each time.
8. Vacuum centrifuge the eluted peptides for ~15 min (time is dependent on exact model of vacuum centrifuge) until ~5 µl total volume remains, then add 5 µl loading buffer and mix the sample well by pipetting up and down. Briefly vortex and spin down for 1 min to collect the entire sample in the bottom of the well.
9. Run 5 µl of each sample on a mass spectrometer with nanospray source according to the manufacturer's instructions.

   For example, run 2-hr gradients on 15-cm columns, and 4-hr gradients on 50-cm columns to gain optimal (phospho-) proteome coverage.

ANALYZING PHOSPHORYLATION DATA AND CONSTRUCTING QUANTITATIVE NETWORK MODELS

BASIC PROTOCOL 5

After generating the MS data, the raw spectra have to be searched against a protein database in order to match them against possible peptides from which the observed proteins and phosphorylation sites can be identified. Several search algorithms exist, some of the most popular being MaxQuant, ProteomeDiscoverer/SEQUEST, and Mascot (Link et al., 1999; Perkins et al., 1999; Cox and Mann, 2008; Cox et al., 2011). While these algorithms can all identify and quantitate peptides and proteins, they have different accuracies and specific requirements, of which an extensive discussion is beyond the

scope of this unit. Here, focus is on MaxQuant, as it is a relatively user-friendly tool that enables custom confidence thresholds to be set, is actively maintained (Cox and Mann, 2008), and is free-of-charge. Moreover, if samples are not labeled, MaxQuant allows one to conduct label-free quantitation. However, this approach suffers from lower accuracy than labeling-based quantitation as it compares peptide abundances between samples, which have been prepared and analyzed separately. Therefore, they are likely to be affected by (slightly) different sample preparation and analysis conditions, which gives rise to artificial experimental artifacts that will influence the data. If possible, one should therefore opt for labeled approaches, but it may nevertheless still prove useful in specific cases where labeled approaches are impossible (Cox and Mann, 2008).

As briefly introduced earlier, an important aim when constructing quantitative phosphorylation networks is to derive crucial kinase-substrate interactions, which may be involved in the phenotype that one is investigating. An increased or decreased activity of one or several kinase(s) involved in this phenotype will likely be manifested in modulated phosphorylation sites, which may show higher or lower abundance. Interpreting phosphorylation site modulation allows for the determination of kinases that are differentially active between experimental conditions. However, a common pitfall that must be taken into account is the importance of distinguishing whether an increase of phosphorylation site abundance is due to the substrate protein having been phosphorylated more (thereby indicating an increased level of kinase activity), or whether the substrate protein was more abundant, thereby explaining the increased levels of the observed phosphopeptide(s). In the case of trying to determine dysregulated kinase-substrate networks, the latter would give rise to false conclusions and should be avoided where possible. This can be controlled for by comparing the phosphorylation levels to the protein levels, which is why it is critical to, in addition to the phosphoproteomic samples, analyze the proteome samples as mentioned in Basic Protocol 1, step 22. This allows for the normalization of the phosphorylation levels to their respective protein abundance, thereby more accurately acting as a proxy for kinase activity (Wu et al., 2011a; see Fig. 11.11.3).

As mentioned above, inferring kinase activity from phosphorylation levels requires computational analyses, which, based on sequence-motif information of the sequence window around a given phosphorylation site combined with the signaling network context of the kinase-substrate interaction, can predict likely kinases to have phosphorylated observed phosphorylation sites. Several approaches have been published over the years, including GPS (Xue et al., 2008), KinasePhos (Wong et al., 2007), NetPhosK (Miller and Blom, 2009), and Scansite (Obenauer et al., 2003), but here a methodology using NetPhorest (Miller et al., 2008) and NetworKIN (Linding et al., 2007), which are developed in-house and have now been combined into a framework known as KinomeXplorer (Horn et al., unpub. observ.), is described. The main reasons for using these two algorithms are (1) they are kept up-to-date on a regular basis, thereby including the latest knowledge in the field, (2) they have been benchmarked intensively to provide their users with accurate modeling capabilities (Miller et al., 2008), (3) they generate probabilities for their predictions, thus allowing probabilistic integration with other types of data and use of confidence thresholds to filter results, and (4) they provide the user with a convenient Web interface, enabling analysis of large datasets in a semi-automated fashion. Additionally, NetPhorest and NetworKIN will generate predictions for other phospho-binding domains interacting with observed phosphorylation sites, enabling more comprehensive modeling to be conducted. Even though these algorithms do not have complete kinome coverage (222 out of 538 at the time of writing), they have the highest coverage compared to alternatives, and additional kinases will be included as the required data becomes available. Below, a computational workflow that allows for the construction of quantitative phosphorylation signaling networks, potentially highlighting kinases of interest in the biological phenomenon that is being investigated, is described. As the above-mentioned

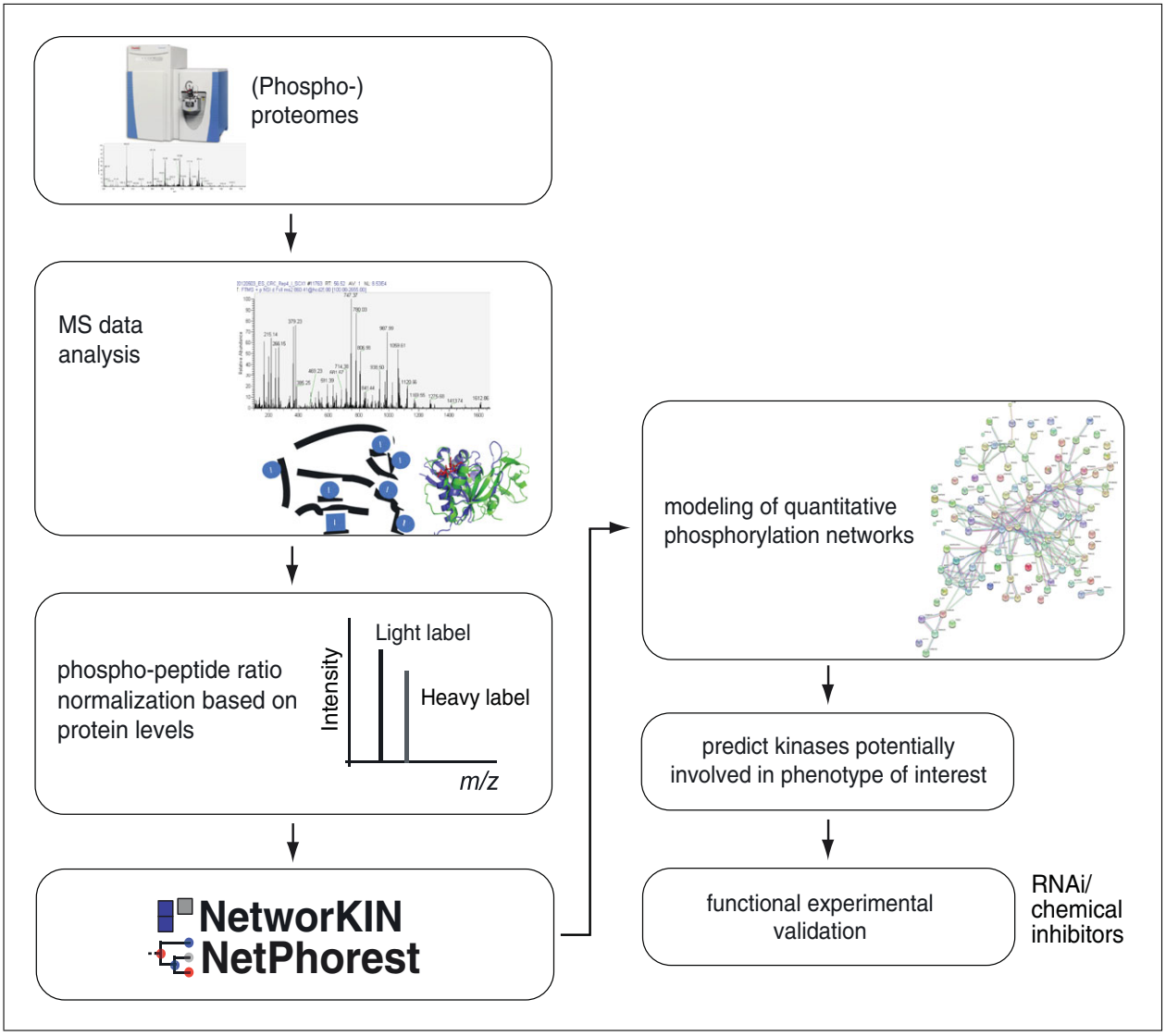


Figure 11.11.2 Modeling workflow overview, detailing the different steps required for constructing quantitative network models that can be used to guide follow-up functional validation in the laboratory.

database-searching software packages provide extensive documentation, here, data analysis steps once the raw MS data has been searched are described and the user has the lists of identified proteins and phosphorylation sites with their corresponding quantitative ratios. See Figure 11.11.2 for an overview of the modeling workflow.

Materials

Desktop computer with Internet access
Mass spectrometry spectral matching software (e.g., MaxQuant or Proteome Discoverer/SEQUEST)
R statistical software
Visual network editing software (e.g., Gephi.org or Cytoscape.org)

1. Conduct a database search for protein and phosphopeptide identification and quantification. In a larger project, it is very useful to rely on a fixed database and release version, e.g., ENSEMBL, to enable easy sequence tracking and mapping. Set the false discovery rate (FDR) to 1% to minimize false-positive protein and phosphorylation site identifications (Elias and Gygi, 2007).

2. Filter out the identified phosphorylation sites with a localization probability <0.75, as for these peptides, the exact position of the phosphorylated residue cannot be assigned with reasonable accuracy (Beausoleil et al., 2006; Taus et al., 2011).

3. Transform the protein/phosphorylation site ratios to $\log_2$. This balances out the positive and negative ratios, as down-regulated proteins/phosphorylation sites would otherwise have ratios between 0 and 1, whereas up-regulated proteins/phosphorylation sites would have ratios from 1 to ∞.

   Log_2 transformation ensures a more accurate and direct comparison between up- and down-regulated peptides.

4. Statistically test protein/phosphorylation site ratios for significance. Using the R statistical software package, use the two-sided, unpaired Mann-Whitney Wilcoxon test. Ratios that have a p value of <0.05 can be considered as significantly up/down modulated, whereas ratios with a p value >0.85 should be considered as being non-modulated (Jorgensen et al., 2009).

5. Where possible, normalize modulated phosphorylation site ratios with their respective parent protein ratios. Parent protein ratios ideally are determined from peptides originating from the same protein that cannot contain a PTM (i.e., peptides without serine, threonine, tyrosine, and methionine residues). If at least three unique peptides are observed for a given protein, its ratio can be determined by taking the mean of all unique peptide ratios. This is to ensure that an observed increase in phosphorylation is due to an increase in kinase activity, rather than an increase in substrate protein. This can be accomplished by dividing the phosphorylation site ratio by the protein ratio, as this normalizes the phosphorylation abundance compared to the protein abundance and filters out any phosphorylation site modulation only due to increased protein abundance or degradation (see Fig. 11.11.3). Additionally, protein phosphorylation stoichiometry should be investigated to gain a better perspective of the phosphorylation dynamics (Wu et al., 2011b). This is again to ensure that observed phosphopeptide regulation is due to altered kinase or phosphatase activity, rather than altered protein expression levels or protein degradation.

6. Once the significantly modulated phosphorylation sites have been accurately determined, the NetworKIN and NetPhorest algorithms can be accessed via the portal KinomeXplorer.info to predict the modulating kinases. For this, it is required to know the protein sequence and the absolute location of the phosphorylation site within the protein, which can be extracted from the database search results. This information can be submitted to the KinomeXplorer Website (*http://www.kinomeXplorer.info*), which will generate all possible predicted kinases for the submitted phosphorylation sites. Due to the probabilistic nature of the framework, confidence filtering of the results can be done and one should only include predictions with a score >1. Additionally, as there will generally be multiple predicted kinases for a particular phosphorylation site, only the top scoring kinase and kinases having a probability within 30% of the top scoring kinase should be included for further analysis.

7. To more accurately model the phosphorylation networks, it must be determined which kinases have been experimentally observed in the MS experiment. This can be achieved by, e.g., using the protein identification lists and a filtering method, either through a scripting language (e.g., Python or Perl) or the VLOOKUP function in Excel. At the time of writing, the KinomeXplorer framework has not included this functionality, but this will be implemented shortly. By filtering the kinase predictions to only include kinases that were experimentally observed in the cell type(s) that was analyzed, more in vivo/in vitro relevance can be extended to the in silico predictions. In cases where the phosphoproteome is sequenced enough (i.e., coverage of a

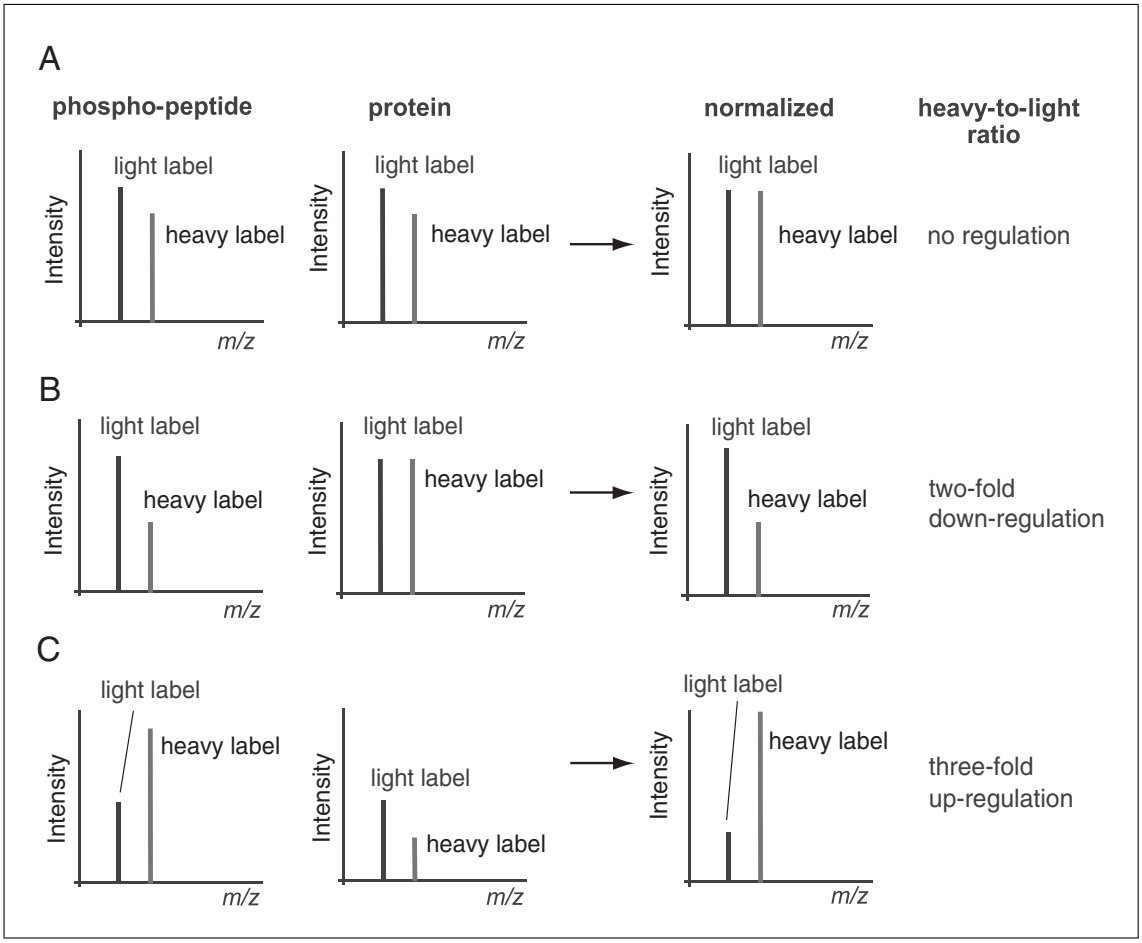


Figure 11.11.3 Phosphorylation site ratio normalization based on protein abundance corrects for protein abundance affecting phosphorylation site abundance, rather than regulated kinase activity. The examples shown are: (**A**) a phosphopeptide that is down-regulated in the heavy labeled sample, whose parent protein is also down-regulated, should be considered as non-regulated. (**B**) A phosphopeptide that is down-regulated in the heavy labeled sample, whose parent protein shows no regulation, should be considered as down-regulated. (**C**) A phosphopeptide that is up-regulated in the heavy sample, whose parent protein is down-regulated, should be considered as an increased up-regulated peptide.

representative subset of the kinome), this principle can be extended to only include kinase predictions from kinases for which a so-called regulatory phosphorylation site has been observed. This is based on the principle that many kinases have a regulatory loop containing a specific residue that is required to be phosphorylated for the kinase to be catalytically activated (or inactivated) (Jorgensen et al., 2009). This will, in the near future, be a built-in function of KinomeXplorer, which will help automate the data processing steps. The regulatory phosphorylation sites, which have been annotated from the literature, can be extracted from public resources such as PhosphoSitePlus (Hornbeck et al., 2012), but it should be noted that deploying this extent of filtering stringency requires considerable depth of kinome coverage in the phosphoproteome data and may not always be feasible. Furthermore, despite on-going efforts, knowledge about these regulatory phosphorylation sites is somewhat limited, so their filtering cannot be applied at a kinome-wide scale.

8. Once the set of kinase predictions has been filtered to include only experimentally supported predicted kinases and their observed substrates, insight into enriched kinase activity and altered signaling networks can be gained. For the former, it can be investigated whether a specific group of kinases is predicted to be more active in one

experimental condition than another, by dividing the total number of phosphorylation sites a particular kinase is predicted for by the total number of phosphorylation sites modulated in the same fashion (up/down). This allows for inter-kinase and inter-experimental comparisons, and can elucidate key kinases, which may display different activity levels. To extend this enrichment analysis to a more global (e.g., disease- or condition-specific) level, enrichment should be calculated compared to kinase enrichment in a large collection of known phosphorylation sites such as phospho.ELM (Dinkel et al., 2011) or PhosphoSitePlus (Hornbeck et al., 2012), as this normalizes experimental kinase activity enrichment to a global activity profile (Van Hoof et al., 2009).

9. For a more visual representation and potential mechanistic insight into the signaling network dynamics, an overview of the kinase-substrate interactions can be obtained by importing the filtered predictions into a visual network editor such as Cytoscape (Shannon et al., 2003) or Gephi (Bastian and Heymann, 2009). Here, specific color coding can be deployed to distinguish between up- and down-regulated kinase-substrate interactions, which can often pinpoint specific kinases that become differentially regulated under given experimental conditions or time-points. If one is mainly interested in kinase-kinase networks, it is useful to draw up the networks of kinases that are predicted to phosphorylate each other, together with the observed substrates they are predicted to phosphorylate. This may allow for the elucidation of a core kinase-substrate network, driven by the interaction of several kinases, which may be involved in the phenotype under investigation.

REAGENTS AND SOLUTIONS

Use deionized, distilled water in all recipes and protocol steps. For common stock solutions, see ***APPENDIX 2A****; for suppliers, see* ***APPENDIX 5****.*

Acidification buffer

1% trifluoroacetic acid (TFA; Sigma, cat. no. T6508)
5% acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
MilliQ H_2O
Store up to 1 week at room temperature

Buffer A

0.1% formic acid, HPLC-grade (Fisher Scientific, cat. no. A117-50)
H_2O, HPLC-grade (Sigma, cat. no. 39253)
Store up to 1 month at room temperature

Buffer B

80% acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
0.1% formic acid, HPLC-grade (Fisher Scientific, cat. no. A117-50)
H_2O, HPLC-grade (Sigma, cat. no. 39253)
Store up to 1 month at room temperature

Denaturation buffer

6 M urea (Sigma, cat. no. 15604)
2 M thiourea (Sigma, cat. no. T7875)
10 mM HEPES, pH 8 (Sigma, cat. no. H4034)
Prepare fresh or store upto 6 months at −80°C
Never heat >25°C

Dimethyl labeling solution (Boersema et al., 2009)

Volumes based on one sample that is to be labeled (adjust as necessary):
4.5 ml 50 mM sodium phosphate buffer, pH 7.5 (mix 1 ml of 50 mM NaH_2PO_4 with 3.5 ml of 50 mM Na_2HPO_4)
250 µl 4% (v/v) formaldehyde in MilliQ H_2O (CH_2O for light, CD_2O for medium, $^{13}CD_2O$ for heavy)
250 µl 0.6 M cyanoborohydride in MilliQ H_2O ($NaBH_3CN$ for light or $NaBD_3CN$ for medium/heavy labels)
Store for maximum 24 hr at 4°C

Formaldehyde (CH_2O) (37% (v/v), Sigma, cat. no. 252549)

Formaldehyde (CD_2O) (20%, 98% D, Isotec, cat. no. 492620)

Formaldehyde ($^{13}CD_2O$) (20%, 99% 13C, 98% D, Isotec, cat. no. 596388)

Sodium cyanoborohydride ($NaBH_3CN$) (Fluka, cat. no. 71435)

Sodium cyanoborodeuteride ($NaBD_3CN$) (96% D, Isotec, cat. no. 190020)

Sodium dihydrogen phosphate (NaH_2PO_4) (Merck, cat. no. 1.06346)

Di-sodium hydrogen phosphate (Na_2HPO_4) (Merck, cat. no. 1.06580)

Loading buffer

1% trifluoroacetic acid (TFA; Sigma, cat. no. T6508)
2% acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
H_2O, HPLC-grade (Sigma, cat. no. 39253)
Store up to 2 weeks at room temperature

Modified RIPA buffer

50 mM Tris·Cl, pH 7.5 (Sigma, cat. no. T3253)
150 mM NaCl (Sigma, cat. no. S7653)
1% NP40/IgePal (Sigma, cat. no. I8896)
0.5% Na-deoxycholate (Sigma, cat. no. D6750)
1 mM EDTA (Sigma, cat. no. E1644)
β-glycerophosphate (5 mM final concentration) (Sigma, cat. no. G9422), add fresh
NaF (5 mM final concentration) (Sigma, cat. no. S7920), add fresh
Na-orthovanadate (activated; Gordon et al., 1991; 1 mM final concentration) (Sigma, cat. no. 450243), add fresh
Roche complete protease inhibitor cocktail (one tablet added fresh per 10 ml RIPA buffer) (Roche, cat. no. 05 892 791 001)
Store up to 6 months at −20°C

Sample buffer

3% acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
1% trifluoroacetic acid (TFA; Sigma, cat. no. T6508)
H_2O, HPLC-grade (Sigma, cat. no. 39253)
Store up to 2 weeks at room temperature

SCX buffer A

5 mM potassium dihydrogen phosphate (Sigma, cat. no. P9791)
30% acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
70% H_2O, HPLC-grade (Sigma, cat. no. 39253)
pH 2.7 with TFA
Store up to 1 month at room temperature

SCX buffer B

5 mM potassium dihydrogen phosphate (Sigma, cat. no. P9791)
350 mM potassium chloride (Millipore, cat. no. 1.04936.0500)
30% acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
70% H_2O, HPLC-grade (Sigma, cat. no. 39253)
pH 2.7 with TFA
Store up to 1 month at room temperature

TiO_2 elution buffer 1

5% ammonia solution (Emsure, cat. no. 1.05432.1000)
H_2O, HPLC-grade (Sigma, cat. no. 39253)
Store up to 3 days at room temperature

TiO_2 elution buffer 2

10% ammonia solution (Emsure, cat. no. 1.05432.1000)
25% acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
H_2O, HPLC-grade (Sigma, cat. no. 39253)
Store up to 3 days at room temperature

TiO_2 loading solution

20 mg/ml 2,5-dihydroxybenzoic acid (Sigma, cat. no. 85707)
5% trifluoroacetic acid (TFA; Sigma, cat. no. T6508)
30% acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
H_2O, HPLC-grade (Sigma, cat. no. 39253)
Store up to 1 month at 4°C

TiO_2 washing solution 1

40% acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
0.25% acetic acid, HPLC-grade (Fisher Scientific, cat. no. A35-500)
0.5% trifluoroacetic acid (TFA; Sigma, cat. no. T6508)
H_2O, HPLC-grade (Sigma, cat. no. 39253)
Prepare fresh

TiO_2 washing solution 2

80% acetonitrile, HPLC-grade (Sigma, cat. no. 34851N)
0.5% acetic acid, HPLC-grade (Fisher Scientific, cat. no. A35-500)
Store up to 1 week at room temperature

COMMENTARY

Background Information

The techniques described in this unit enable a biological system under investigation to be modeled from the phosphorylation-based signaling perspective, potentially highlighting key proteins involved in a phenotype of interest. By utilizing experimental data as input for computational modeling, more in-depth insight about the signaling networks can be obtained, the results of which can be used to drive follow-up validation studies. In any data analysis approach involving computational predictions, it is of vital importance to experimentally validate (some of) the predictions, as this helps ensure that the predictions are biologically relevant and helps to guide threshold settings. Any key kinases determined in Basic Protocol 5, steps 8 and 9, should be used as input for guiding subsequent experimental validation studies, where the exact role of these kinases in a given phenotype should be functionally assessed by, e.g., RNAi or chemical inhibitor experiments. This can give conclusive evidence of whether or not a kinase or group of kinases are required for a specific phenotype, disease progression, or drug resistance development (Bakal et al., 2008; Jorgensen et al., 2009; Lee et al., 2012). Preferably, this is also done in a time-staggered manner

to monitor the cellular responses to a perturbation/stimulation or combination thereof, as this will elucidate a more complete picture of the altered signaling dynamics within the cell and enables one to tweak the resulting model to higher accuracy. In complex diseases such as cancer, but also in immune response–dependent signaling, this can give more insight into potential treatment strategies, as better understanding of the signaling networks is obtained. By integrating computational and experimental approaches, the strengths of both techniques can be combined, facilitating some limitations of either technique to be (partially) overcome. Finally, the modeling capacities generated by the KinomeXplorer (and underlying NetworKIN and NetPhorest algorithms) framework will grow with the availability of additional kinase-substrate recognition and kinase-substrate interaction data, enabling one to extend established kinase-substrate network models to a kinome-wide level.

Critical Parameters

The most critical parameter of the methods described include conducting different experimental steps in a swift yet cautious manner, due to the labile nature of phosphorylated peptides. Stationary waiting stages should be kept to a minimum, as it is critical to have the enriched samples analyzed as quickly as possible. Additionally, due to the moderately volatile nature of many of the buffers (mainly acetonitrile- and ammonia-containing ones), preparing fresh buffers is imperative and they should be replaced as indicated. Additionally, all of the reagents utilized should be of HPLC quality to reduce contamination of the instrument, and likewise gloves should be worn throughout the protocol to minimize keratin contamination of the sample.

Troubleshooting

In the case of inadequate quantitative data generated, it should be investigated whether this could be attributed to inefficient labeling. The simplest way of checking for this is to run a small aliquot of the labeled samples individually, and to search for unlabeled peptides (Boersema et al., 2009). If this is the case, repeat Basic Protocol 1 until full labeling is achieved. In the case of low phosphoproteome coverage, several possible causes can be identified, and pinpointing the exact one(s) becomes a challenge. Generally, it is vital to ensure a quick lysis procedure with ice-cold buffers and adequate protease and phosphatase inhibitors as described above, to ensure the preservation of the phosphorylated proteins. Additionally, it is important to monitor pH levels as indicated in the protocol, and to ensure that vacuum centrifugation is done correctly to eliminate organic solvents in the sample. Finally, in the case of lack of specificity during the enrichment (i.e., a large number of unphosphorylated peptides being detected), make sure the washing steps are carried out accurately, removing as much of the supernatant as possible without disturbing the pellet.

Anticipated Results

Depending on the amount of starting material, the number of unique phosphorylation sites that can be identified should range between hundreds and tens of thousands. Using this protocol in house, the authors identify between ~1000 to 4000 unique phosphorylation sites with <2 mg of starting material without SCX fractionation, and ~20,000 phosphorylation sites with 24 mg of starting material. Results will vary, however, depending mainly on the biological system under investigation, instrument performance, and experience.

Time Considerations

Lysing of cells requires 1 hr and acetone precipitation should be done overnight. Dissolving protein in denaturation buffer requires between a few hours and an overnight incubation. Reduction and alkylation require 1 day and an overnight digestion. Dimethyl labeling, SCX fractionation, and phospho-enrichment require 1 day. Mass spectrometry depends on the number of samples and gradient times. Data analysis is dependent on the number of samples and computer performance; it will require anywhere between days and weeks.

Acknowledgments

The authors thank Pau Creixell (C-SIG, DTU) for comments on the manuscript, Chiara Francavilla and Jesper Olsen (CPR, KU) for input on protocols, and Jonas Nørskov Søndergaard and Susanne Brix Pedersen (CBS, DTU) for providing primary cell material. This work was supported by the Lundbeck Foundation, the Human Frontier Science Program (HFSP), the Danish Council for Independent Research (FSS), and the European Research Council (ERC). Visit *http://www.networkbio.org/*, *http://www.lindinglab.org/* for more information on cancer-related network biology.

Literature Cited

Alpert, A.J. 2008. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem.* 80:62-76.

Bakal, C., Linding, R., Llense, F., Heffern, E., Martin-Blanco, E., Pawson, T., and Perrimon, N. 2008. Phosphorylation networks regulating JNK activity in diverse genetic backgrounds. *Science* 322:453-456.

Bastian, M. and Heymann, S. 2009. Gephi: An open source software for exploring and manipulating network. International AAAI Conference on Weblogs and Social Media. *https://gephi.org*

Beausoleil, S.A., Villén, J., Gerber, S.A., Rush, J., and Gygi, S.P. 2006. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* 24:1285-1292.

Bodenmiller, B. and Aebersold, R. 2010. Quantitative analysis of protein phosphorylation on a system-wide scale by mass spectrometry-based proteomics. *Methods Enzymol.* 470:317-334.

Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A.J. 2009. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* 4:484-494.

Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72:248-254.

Brognard, J. and Hunter, T. 2011. Protein kinase signaling networks in cancer. *Curr. Opin. Genet. Dev.* 21:4-11.

Burckstummer, T., Bennett, K.L., Preradovic, A., Schütze, G., Hantschel, O., Superti-Furga, G., and Bauch, A. 2006. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat. Methods* 3:1013-1019.

Callister, S.J., Barry, R.C., Adkins, J.N., Johnson, E.T., Qian, W.J., Webb-Robertson, B.J., Smith, R.D., and Lipton, M.S. 2006. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* 5:277-286.

Cannons, J.L. and Schwartzberg, P.L. 2004. Fine-tuning lymphocyte regulation: What's new with tyrosine kinases and phosphatases? *Curr. Opin. Immunol.* 16:296-303.

Cantley, L.C., Auger, K.R., Carpenter, C., Duckworth, B., Graziani, A., Kapeller, R., and Soltoff, S. 1991. Oncogenes and signal transduction. *Cell* 64:281-302.

Cox, J. and Mann, M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26:1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. 2011. Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10:1794-1805.

Creixell, P., Schoof, E.M., Erler, J.T., and Linding, R. 2012. Navigating cancer network attractors for tumor-specific therapy. *Nat. Biotechnol.* 30:842-848.

Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J., and Diella, F. 2011. Phospho.ELM: A database of phosphorylation sites—Update 2011. *Nucleic Acids Res.* 39:D261-D267.

Dustin, M.L. 2009. The cellular context of T cell signaling. *Immunity* 30:482-492.

Dyson, M.R., Zheng, Y., Zhang, C., Colwill, K., Pershad, K., Kay, B.K., Pawson, T., and McCafferty, J. 2011. Mapping protein interactions by combining antibody affinity maturation and mass spectrometry. *Anal. Biochem.* 417:25-35.

Elias, J.E. and Gygi, S.P. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4:207-214.

Engholm-Keller, K., Birck, P., Størling, J., Pociot, F., Mandrup-Poulsen, T., and Larsen, M.R. 2012. TiSH—A robust and sensitive global phosphoproteomics strategy employing a combination of TiO(2), SIMAC, and HILIC. *J. Proteomics* 75:5749-5761.

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

Fedorov, O., Muller, S., and Knapp, S. 2010. The (un)targeted cancer kinome. *Nat. Chem. Biol.* 6:166-169.

Gordon, J.A. 1991. Use of vanadate as protein-phosphotyrosine phosphatase inhibitor. *Methods Enzymol.* 201:477-482.

Hjerrild, M., Stensballe, A., Rasmussen, T.E., Kofoed, C.B., Blom, N., Sicheritz-Ponten, T., Larsen, M.R., Brunak, S., Jensen, O.N., and Gammeltoft, S. 2004. Identification of phosphorylation sites in protein kinase A substrates using artificial neural networks and mass spectrometry. *J. Proteome Res.* 3:426-433.

Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. 2012. PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* 40:D261-D270.

Hörth, P., Miller, C.A., Preckel, T., and Wenz, C. 2006. Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol. Cell. Proteomics* 5:1968-1974.

Isakov, N. and Altman, A. 2002. Protein kinase C(theta) in T cell activation. *Annu. Rev. Immunol.* 20:761-794.

Janes, K.A., Albeck, J.G., Gaudet, S., Sorger, P.K., Lauffenburger, D.A., and Yaffe, M.B. 2005. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* 310:1646-1653.

Jensen, K.J. and Janes, K.A. 2012. Modeling the latent dimensions of multivariate signaling datasets. *Phys. Biol.* 9:045004.

Jiang, Z.T. and Zuo, Y.M. 2001. Synthesis of porous titania microspheres for HPLC packings by polymerization-induced colloid aggregation (PICA). *Anal. Chem.* 73:686-688.

Jin, L.L., Tong, J., Prakash, A., Peterman, S.M., St-Germain, J.R., Taylor, P., Trudel, S., and Moran, M.F. 2010. Measurement of protein phosphorylation stoichiometry by selected reaction monitoring mass spectrometry. *J. Proteome Res.* 9:2752-2761.

Jorgensen, C., Sherman, A., Chen, G.I., Pasculescu, A., Poliakov, A., Hsiung, M., Larsen, B., Wilkinson, D.G., Linding, R., and Pawson, T. 2009. Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells. *Science* 326:1502-1509.

Kawahara, M., Nakamura, H., and Nakajima, T. 1990. Titania and zirconia: Possible new ceramic microparticulates for high-performance liquid chromatography. *J. Chromatogr. A* 515:149-158.

Kreeger, P.K., Wang, Y., Haigis, K.M., and Lauffenburger, D.A. 2010. Integration of multiple signaling pathway activities resolves K-RAS/N-RAS mutation paradox in colon epithelial cell response to inflammatory cytokine stimulation. *Integr. Biol. (Camb.)* 2:202-208.

Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P., and Jørgensen, T.J. 2005. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* 4:873-886.

Lee, M.J., Ye, A.S., Gardino, A.K., Heijink, A.M., Sorger, P.K., MacBeath, G., and Yaffe, M.B. 2012. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* 149:780-794.

Lemmon, M.A. and Schlessinger, J. 2010. Cell signaling by receptor tyrosine kinases. *Cell* 141:1117-1134.

Linding, R. 2010. Multivariate signal integration. *Nat. Rev. Mol. Cell. Biol.* 11:391.

Linding, R., Jensen, L.J., Ostheimer, G.J., van Vugt, M.A., Jørgensen, C., Miron, I.M., Diella, F., Colwill, K., Taylor, L., Elder, K., Metalnikov, P., Nguyen, V., Pasculescu, A., Jin, J., Park, J.G., Samson, L.D., Woodgett, J.R., Russell, R.B., Bork, P., Yaffe, M.B., and Pawson, T. 2007. Systematic discovery of in vivo phosphorylation networks. *Cell* 129:1415-1426.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R. 3rd. 1999. Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* 17:676-682.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. 2002. The protein kinase complement of the human genome. *Science* 298:1912-1934.

McNulty, D.E. and Annan, R.S. 2008. Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol. Cell. Proteomics* 7:971-980.

Michel, P.E., Reymond, F., Arnaud, I.L., Josserand, J., Girault, H.H., and Rossier, J.S. 2003. Protein fractionation in a multicompartment device using off-gel isoelectric focusing. *Electrophoresis* 24:3-11.

Miller, A.T. and Berg, L.J. 2002. New insights into the regulation and functions of Tec family tyrosine kinases in the immune system. *Curr. Opin. Immunol.* 14:331-340.

Miller, M.L. and Blom, N. 2009. Kinase-specific prediction of protein phosphorylation sites. *Methods Mol. Biol.* 527:299-310.

Miller, M.L., Jensen, L.J., Diella, F., Jørgensen, C., Tinti, M., Li, L., Hsiung, M., Parker, S.A., Bordeaux, J., Sicheritz-Ponten, T., Olhovsky, M., Pasculescu, A., Alexander, J., Knapp, S., Blom, N., Bork, P., Li, S., Cesareni, G., Pawson, T., Turk, B.E., Yaffe, M.B., Brunak, S., and Linding, R. 2008. Linear motif atlas for phosphorylation-dependent signaling. *Sci. Signal.* 1:ra2.

Miller-Jensen, K., Janes, K.A., Brugge, J.S., and Lauffenburger, D.A. 2007. Common effector processing mediates cell-specific responses to stimuli. *Nature* 448:604-608.

Mohammed, S. and Heck, A. Jr. 2011. Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. *Curr. Opin. Biotechnol.* 22:9-16.

Monetti, M., Nagaraj, N., Sharma, K., and Mann, M. 2011. Large-scale phosphosite quantification in tissues by a spike-in SILAC method. *Nat. Methods* 8:655-658.

Monks, C.R., Freiberg, B.A., Kupfer, H., Sciaky, N., and Kupfer, A. 1998. Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature* 395:82-86.

Munoz, J. and Heck, A.J. 2011. Quantitative proteome and phosphoproteome analysis of human pluripotent stem cells. *Methods Mol. Biol.* 767:297-312.

Obenauer, J.C., Cantley, L.C., and Yaffe, M.B. 2003. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* 31:3635-3641.

Olsen, J.V. and Macek, B. 2009. High accuracy mass spectrometry in large-scale analysis of protein phosphorylation. *Methods Mol. Biol.* 492:131-142.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127:635-648.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. 2002. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1:376-386.

Pandey, A., Podtelejnikov, A.V., Blagoev, B., Bustelo, X.R., Mann, M., and Lodish, H.F. 2000. Analysis of receptor signaling pathways by mass spectrometry: Identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc. Natl. Acad. Sci. U.S.A.* 97:179-184.

Pawson, T. 1995. Protein modules and signaling networks. *Nature* 373:573-580.

Pawson, T. and Hunter, T. 1994. Signal transduction and growth control in normal and cancer cells. *Curr. Opin. Genet. Dev.* 4:1-4.

Pawson, T. and Kofler, M. 2009. Kinome signaling through regulated protein-protein interactions in normal and cancer cells. *Curr. Opin. Cell. Biol.* 21:147-153.

Pawson, T. and Linding, R. 2008. Network medicine. *FEBS Lett.* 582:1266-1270.

Pedersen, M.W., Jacobsen, H.J., Koefoed, K., Hey, A., Pyke, C., Haurum, J.S., and Kragh, M. 2010. Sym004: A novel synergistic anti-epidermal growth factor receptor antibody mixture with superior anticancer efficacy. *Cancer Res.* 70:588-597.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551-3567.

Posewitz, M.C. and Tempst, P. 1999. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* 71:2883-2892.

Prakash, A., Piening, B., Whiteaker, J., Zhang, H., Shaffer, S.A., Martin, D., Hohmann, L., Cooke, K., Olson, J.M., Hansen, S., Flory, M.R., Lee, H., Watts, J., Goodlett, D.R., Aebersold, R., Paulovich, A., and Schwikowski, B. 2007. Assessing bias in experiment design for large scale mass spectrometry-based quantitative proteomics. *Mol. Cell. Proteomics* 6:1741-1748.

Rappsilber, J., Mann, M., and Ishihama, Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2:1896-1906.

Readinger, J.A., Mueller, K.L., Venegas, A.M., Horai, R., and Schwartzberg, P.L. 2009. Tec kinases regulate T-lymphocyte development and function: New insights into the roles of Itk and Rlk/Txk. *Immunol. Rev.* 228:93-114.

Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., Hu, Y., Tan, Z., Stokes, M., Sullivan, L., Mitchell, J., Wetzel, R., Macneill, J., Ren, J.M., Yuan, J., Bakalarski, C.E., Villen, J., Kornhauser, J.M., Smith, B., Li, D., Zhou, X., Gygi, S.P., Gu, T.L., Polakiewicz, R.D., Rush, J., and Comb, M.J. 2007. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* 131:1190-1203.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D.J. 2004. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3:1154-1169.

Saez-Rodriguez, J., Alexopoulos, L.G., Epperlein, J., Samaga, R., Lauffenburger, D.A., Klamt, S., and Sorger, P.K. 2009. Discrete logic modelling as a means to link protein signaling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.* 5:331.

Samelson, L.E., Patel, M.D., Weissman, A.M., Harford, J.B., and Klausner, R.D. 1986. Antigen activation of murine T cells induces tyrosine phosphorylation of a polypeptide associated with the T cell antigen receptor. *Cell* 46:1083-1090.

Schirle, M., Heurtier, M.A., and Kuster, B. 2003. Profiling core proteomes of human cell lines by one-dimensional PAGE and liquid chromatography-tandem mass spectrometry. *Mol. Cell. Proteomics* 2:1297-1305.

Seet, B.T., Dikic, I., Zhou, M.M., and Pawson, T. 2006. Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell. Biol.* 7:473-483.

Shah, K. and Shokat, K.M. 2003. A chemical genetic approach for the identification of direct substrates of protein kinases. *Methods Mol. Biol.* 233:253-271.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13:2498-2504.

Shawver, L.K., Slamon, D., and Ullrich, A. 2002. Smart drugs: Tyrosine kinase inhibitors in cancer therapy. *Cancer Cell* 1:117-123.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L.J., and von Mering, C. 2011. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39:D561-D568.

Tan, C.S., Bodenmiller, B., Pasculescu, A., Jovanovic, M., Hengartner, M.O., Jørgensen, C., Bader, G.D., Aebersold, R., Pawson, T., and Linding, R. 2009. Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci. Signal.* 2:ra39.

Tani, K. and Suzuki, Y. 1994. Syntheses of spherical silica and titania from alkoxides on a laboratory scale. *Chromatographia* 38:291-294.

Tanl, K., Sumizawa, T., Watanabe, M., Tachibana, M. Koizumi, H., and Kiba, T. 2002. Evaluation of titania as an ion-exchanger and as a ligand-exchanger in HPLC. *Chromatographia* 55:33-37.

Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. 2011. Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* 10:5354-5362.

Thingholm, T.E., Jørgensen, T.J., Jensen, O.N., and Larsen, M.R. 2006. Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nat. Protoc.* 1:1929-1935.

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. 2003. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75:1895-1904.

Van Hoof, D., Muñoz, J., Braam, S.R., Pinkse, M.W., Linding, R., Heck, A.J., Mummery, C.L., and Krijgsveld, J. 2009. Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell* 5:214-226.

Wong, Y.H., Lee, T.Y., Liang, H.K., Huang, C.M., Wang, T.Y., Yang, Y.H., Chu, C.H., Huang, H.D., Ko, M.T., and Hwang, J.K. 2007. KinasePhos 2.0: A web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res.* 35:W588-W594.

Wu, R., Dephoure, N., Haas, W., Huttlin, E.L., Zhai, B., Sowa, M.E., and Gygi, S.P. 2011a. Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes. *Mol. Cell. Proteomics* 10:M111.009654.

Wu, R., Haas, W., Dephoure, N., Huttlin, E.L., Zhai, B., Sowa, M.E., and Gygi, S.P. 2011b. A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat. Methods* 8:677-683.

Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L., and Yao, X. 2008. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol. Cell. Proteomics* 7:1598-1608.

Zhou, H., Ye, M., Dong, J., Corradini, E., Cristobal, A., Heck, A.J.R., Zou, H., and Mohammed, S. 2013. Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat. Protoc.* 8:461-480.

Internet Resources

http://www.kinomexplorer.info

Internet portal to access the integrated NetPhorest and NetworKIN frameworks.

http://www.networkin.info

Internet portal to access the original NetworKIN framework.

http://www.netphorest.info

Internet portal to access the original NetPhorest framework.

Chapter II

Part III

KinomeXplorer: An Integrated Platform for Kinome Biology Studies

KinomeXplorer: An Integrated Platform for Kinome Biology Studies

Heiko Horn[1*] Erwin M. Schoof[2*], Jinho Kim[2*], Xavier Robin[2], Martin L. Miller[3], Francesca Diella[4,5], Anita Palma[6], Gianni Cesareni[6,7], Rune Linding[2@], Lars Juhl Jensen[1@]

[1] Novo Nordisk Foundation Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen, Denmark

[2] Cellular Signal Integration Group (C-SIG), Center for Biological Sequence Analysis (CBS), Department of Systems Biology, Technical University of Denmark (DTU), Building 301, DK-2800, Lyngby, Denmark

[3] Computational Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

[4] Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

[5] Molecular Health GmbH, Heidelberg, Germany

[6] Department of Biology, University of Rome Tor Vergata, I-00133 Rome, Italy

[7] Istituto Ricovero e Cura a Carattere Scientifico, Fondazione Santa Lucia, Via Ardeatina, 306, 00179 Rome, Italy

* These authors contributed equally to this work.

@ To whom correspondence should be addressed; E-mail: lars.juhl.jensen@cpr.ku.dk and linding@cbs.dtu.dk

Network biology studies of the human kinome are critical for systems-based pharmacology. Here we present KinomeXplorer, an integrated platform for modeling kinome signaling networks by combining sequence specificity with cellular context. This novel framework features a next-generation scoring scheme, resulting in broader coverage, a boost in accuracy and greater usability. Both a local package and an interactive web interface allows investigation of predicted kinase–substrate interactions from human and yeast, as well as modeling human phosphatase–substrate interactions.

Powerful tools for the computational analysis and integration of complex biological data are increasingly lagging behind our ability to generate data. This is true across diverse technologies such as cell imaging, DNA sequencing and mass spectrometry and this in turn creates a bottleneck for predictive modeling of biological systems. Specifically, it has become routine to create large-scale phospho-proteomics datasets to elucidate the phosphorylation events associated with a given phenotype or disease condition such as cancer. In order to fully utilize the power of these large-scale datasets, tools are required to de-convolute the underlying intracellular signaling networks that mediate and respond to these phosphorylation events.

The residues phosphorylated by kinases not only have a direct effect on the substrate protein activity, they also create binding sites for modular phospho-binding domains, thereby giving rise to directionality and logic gating in cellular signaling networks[1,2]. Kinases and phospho-binding proteins typically interact with phosphorylation sites in a transient manner, making these interactions challenging or even impossible to be captured by cellular/in vivo experiments alone. Furthermore, it is difficult to design kinase perturbation experiments because the kinome-wide selectivity and specificity of many kinase inhibitors is unknown[3,4]. As a result, we lack knowledge on which of the approximately 540 human kinases phosphorylate a given site: of the 42,914 phosphorylation sites currently annotated in the

Phospho.ELM database[5], only ~20% have been linked to a kinase. This proportion is swiftly decreasing as technological advances in mass-spectrometry-based phospho-proteomics accelerate our ability to identify phosphorylation sites, but not to determine which kinase(s) phosphorylate(s) them. In other words, we keep discovering new regulatory players, but are unable to place them into the cellular signaling network at the same rate.

To systematically identify these dynamic interactions, computational methods must be deployed to guide experiments. We have shown that combining computer algorithms with quantitative mass-spectrometry is a powerful approach to validate kinase-substrate relationships[6]. Critically, we showed that kinase specificity can be described in terms of two main contributing elements: namely the recognition motif of the individual kinase (e.g. -S/TQ- for the ATM kinase) and proteins that can be functionally associated with it (i.e. not just proteins that directly interact with the kinase). Thus, we previously developed two algorithms to predict kinases for experimentally identified phosphorylation sites, NetPhorest[7] and NetworKIN[6], which are extensively used by researchers worldwide[8-16]. NetPhorest analyses experimentally identified phosphorylation sites, and classifies them according to the linear motif (peptide specificity) of the kinases and phospho-binding domains. Consensus motifs have been shown to be useful in the development of kinase- or kinase-family- specific antibodies, or to detect biases arising from the enrichment procedures commonly used in phospho-proteomics, such as phospho-specific antibodies, IMAC or titanium dioxide (TiO2)[17,18]. Complementary to NetPhorest, NetworKIN models kinase-substrate interactions in an integrative manner by combining sequence specificity motifs from NetPhorest with contextual network information (e.g. protein-protein interactions) from STRING[19]. This database provides the contextual information we utilize to disambiguate between kinases sharing the same or a similar motif (and thus are grouped into one NetPhorest classifier), in addition to adding more confidence to the NetPhorest predictions. The network context of kinases is critical, as exemplified by the discovery that the phenotypic role of JNK kinase depends entirely on the state of the cellular signaling networks prior to its activation[20]. In other words, it is critical to assess the protein networks embedding kinases and how these are dynamically modulated (e.g. through time or perturbations) in order to predict cell behavior[21].

Here we present an integrated platform, KinomeXplorer (Fig. 1), which provides workflows that enable researchers to efficiently analyze phosphorylation-dependent interaction networks and aids them in designing follow-up perturbation experiments. The platform provides the next generations of NetworKIN and NetPhorest, conferring increased prediction accuracy through a completely novel Bayesian schoring scheme, broader kinome coverage, new phosphatome coverage, and a completely re-designed unifying web interface. KinomeXplorer distinguishes itself from other network modeling tools such as ARACNE[22] or SteinerNet[23] by 1) using proteomics data rather than microarray expression data, and 2) by attempting to construct novel kinase-substrate interactions based on experimentally observed phosphorylation sites. To facilitate down-stream interpretation of generated predictions, they can be directly imported into network visualization and interrogation tools such as Cytoscape[24]. Lastly, the framework also integrates the new KinomeSelector tool that helps the user select an optimal kinase panel in order to functionally perturb the predicted phosphorylation signaling networks.

NetPhorest was designed to incorporate new training data as soon as it becomes available[7], and we have thus retrained and benchmarked the algorithm with new experimental data in the form of position-specific scoring matrices (PSSM) and phosphorylation sites for neural network training from the latest Phospho.ELM[5] and PhosphositePlus database releases[25]. This has allowed us to expand the coverage from 179 human protein kinases to 222. We also added 22 human phosphatases, which covers around 60% of human tyrosine phosphatases. Additionally, the kinome-wide sequence specificity matrices for all kinases in the budding yeast (*Saccharomyces cerevisiae*) have been included[26]. The yeast version covers 71 of the 122 protein kinases, giving rise to a ~60% kinome-coverage. As only 111 kinases have been isolated as active enzymes in vitro, this resulted in a de-facto coverage of 64%. Because the NetworKIN algorithm builds upon NetPhorest, the improvements described above have also expanded NetworKIN's coverage to >40% of the human kinome. Additionally, due to the high sequence homology, the human versions of NetworKIN and NetPhorest can also be deployed on mouse data[12,27]. We reengineered the NetworKIN algorithm to further improve the performance and usability. To calculate the NetworKIN score, we combine the NetPhorest probability and the STRING-derived proximity score using the Naive Bayes method, which dramatically improved the prediction accuracy comparing to use of NetPhorest alone (Supplementary Fig 1 and Table 1). By adopting more sophisticated parameters to penalize long paths, NetworKIN is now more reflective of cellular context. We also tackled a well-known but neglected problem in network biology that over-studied proteins cause biases to the network structure. To avoid this bias, we systematically penalize for the connectivity of the intermediated nodes when calculating the network-derived proximity matrix. This novel scoring system brings a significant increase in accuracy ($p < 10^{15}$ over both NetPhorest and the original NetworKIN algorithm), which originates not only from adding more data, but also from completely re-writing the code from scratch and implementing a new statistical framework (Supplementary Fig 1 and Table 1). In terms of usability, the new scoring scheme generates scores to represent how likely the phosphorylation interaction occurs in a probabilistic manner, which facilitates the interpretation of the results. Specifically, predictions with a score higher than the theoretically neutral value of 1 are likely to be true, and the higher the score for a given kinase, the higher the likelihood it is indeed this kinase. This makes it possible to directly compare predictions from different kinases by enabling a single cutoff for all kinases without normalization, which is critical when modeling global kinome networks. Overall this has resulted in a more powerful, accurate and intuitive scoring system which is easier to interpret and deploy, also for wet lab biologists.

KinomeXplorer provides a landing page with a flow-type guide to the user of which sub-module to use for the specific task at hand. For example, if the user is interested in kinase-substrate predictions, it directs the user to the NetworKIN submodule. Furthermore, we have completely redesigned the web interface for both NetPhorest and NetworKIN to make them more intuitive, tightly integrated, and applicable for analyses of large-scale experimental phosphorylation data. The workflow of the new unified interface is highlighted in Supplementary Fig. 2 and text.

Finally, to guide the design of follow-up kinase perturbation studies, we have integrated the KinomeSelector resource (http://kinomeselector.jensenlab.org/) into KinomeXplorer, which provides an interface to construct kinase-panels for inhibition studies. The interface consists of a tree of the human kinome from which the user can select kinases to include in the panel. Additional kinases that fall within a user-customizable threshold are highlighted.

These additional kinases are considered likely targets of the same inhibitors and thus need not be screened separately. The final set of selected kinases and kinases similar to them can be downloaded. We also provide a pre-computed panel of 100 kinase inhibitors (covering 88% of the human kinome).

The structure and dynamics of the cellular signaling networks that control cell behavior are to a large extent determined by the combined actions of kinases, phosphatases and phospho-binding domains. While we can readily assess dynamics of phosphorylation sites, our ability to model and predict the associated networks is a critical area to advance. The importance of this is underlined by the fact that kinases are the target of about 75% of current world-wide drug development programs[28,29] for complex diseases, and it is increasingly evident that they must be targeted in combinations, as elucidated by network models[30]. Therefore, the improvements reported here on the new integrated framework have attempted to further improve the framework as a crucial tool to monitor and model the networks of kinases and their substrates. The resource is shared with both industry and academic research communities and hosted at http://KinomeXplorer.info.

Acknowledgements

We would like to thank members of the Linding Lab for useful input on the manuscript, and Adrian Pasculescu for input on the data collection. This work was supported by the Lundbeck Foundation, the Human Frontier Science Program (HFSP), the Danish Council for Independent Research (FSS) and the European Research Council (ERC).

Author Contributions

R.L. and L.J.J. conceived the project. H.H., E.M.S., J.K., R.L. and L.J.J. designed the experiments. H.H., E.M.S., J.K., X.R., M.L.M, F.D. and G.C. analyzed the data. H.H., J.K and X.R. implemented the web interface. E.M.S., J.K. and M.L.M. updated the NetPhorest framework. A.P., G.C. and F.D. contributed data and phylogenetic trees. H.H., E.M.S., J.K., M.L.M., R.L. and L.J.J. wrote the paper. R.L. and L.J.J. oversaw the project.

Figure legends

Figure1. (a) Score calculation scheme of the NetworKIN algorithm. The NetworKIN algorithm combines network proximity scores and NetPhorest probabilities based on network distances and peptide sequences respectively. The network proximity score is calculated by multiplying the confidence score of each edge while penalizing for the length of the path and the connectivity of intermediate nodes. NetPhorest probabilities are calculated using the trained kinase classifiers, based on the peptide sequences surrounding the phosphorylation site. Then the network proximity scores and NetPhorest probability are converted to likelihood ratios. These two likelihood ratios are combined to generate a unified likelihood ratio. (b) Penalty scheme for hub nodes and path length. Hubs are penalized proportional to their summed up connectivity, based on the confidence scores. Long paths are penalized by multiplying each edge with a correction factor, leading to an exponential correlation between length and final penalty. Hub and length penalty parameters are systematically determined in the NetworKIN benchmarking process. Line widths are proportional to the confidence score. (c) Conversion of Network proximity score and NetPorest probability to likelihood ratios. The likelihood conversion processes are conducted in a kinase specific manner. For each kinase,

data points from positive and negative training sets are collected and arranged by the network proximity score and NetPhorest probability. Then a sliding window along the scores is used to calculate the likelihood ratios.

Supplementary Figure 1. Performance comparison of NetPhorest and NetworKIN (old and new algorithms, highlighting the benefits of adding contextual information. Kinases that have a large enough validation set (at least 20 positive sites) and where STRING could contribute context were taken into consideration. When estimating the NetPhorest performance, we counted a prediction as true if the kinase was part of the kinase family classifier from NetPhorest. The NetworKIN results show that, by adding contextual information, we are able to disambiguate between the potential kinases of one NetPhorest group without losing predictive performance. To evaluate only the effect of the new scoring algorithm, the same NetPhorest engine and benchmark data set were used. It is important to note that while the improvement in AUC may appear modest, the corresponding improvement in practical accuracy is typically several fold.

Supplementary Figure 2. The workflow of the KinomeXplorer web interface. (a) NetworKIN accepts sequences, Gene/Protein names or MaxQuant/ProteomeDiscoverer results as input. In case of names, the Reflect web-service is used to find the best matches in Ensembl v59 and the corresponding sequence will be taken. (b) Sites can be selected in multiple ways: 1) manually by clicking, 2) based on a predictor for phosphorylation probability, 3) known high/low-throughput sites extracted from KinomeXplorer-DB and 4) by uploading a file with a list of sites. Predictions will only be made for the selected sites. (c) In the result page, multiple filtering possibilities are given: 1) filtering by minimum score, 2) filtering by maximum distance from the best scoring kinase for a given site, 3) filtering by tree (user must select which should be shown). For performance reasons, only the best five predictions are shown by default. (d) After having filtered the results, the user can download the result as displayed or select to retrieve the full set of predictions.

Supplementary Table 1. Benchmark of the NetworKIN method. Kinases/phospho-binding domains with more than or equal to 10 known phosphosites are shown.

Supplementary Table 2. Dataset for training NetPhorest and NetworKIN.

Figures

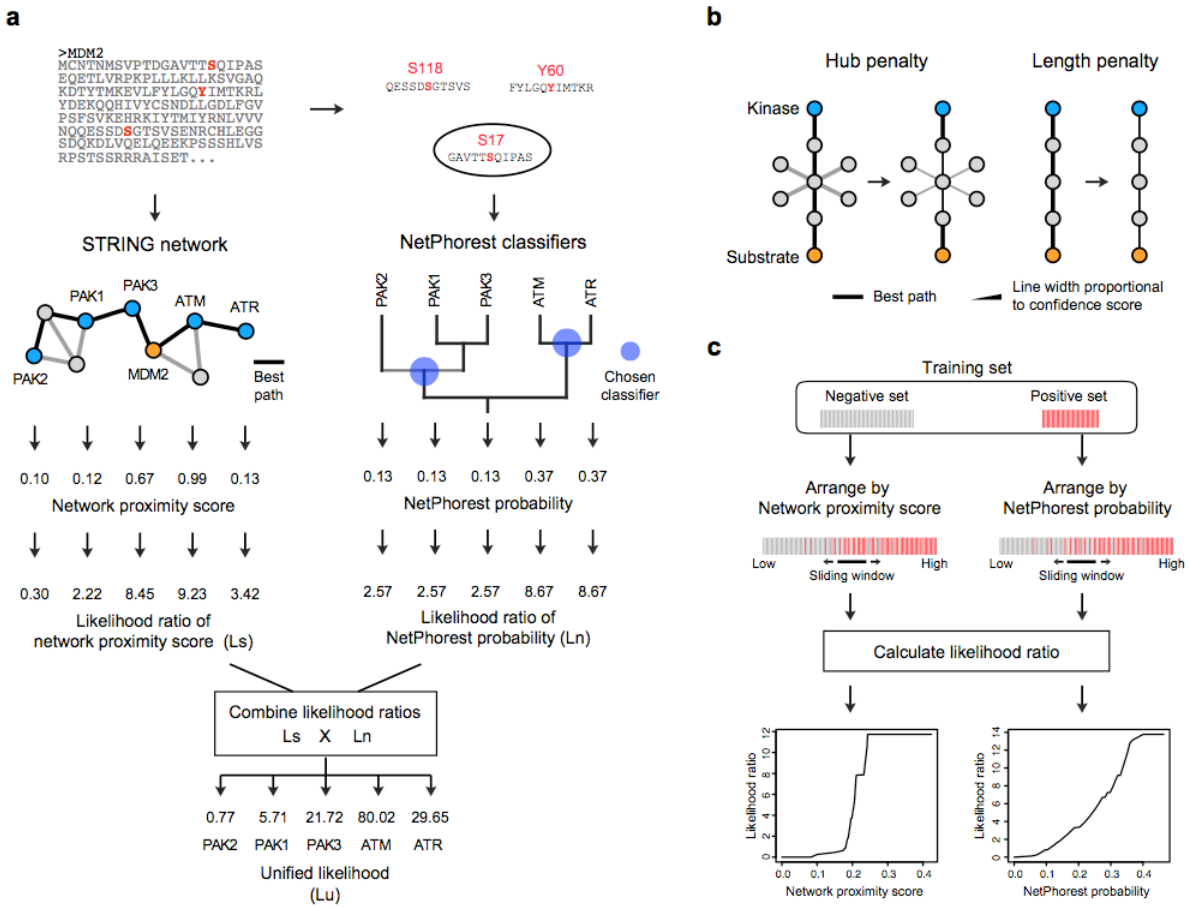
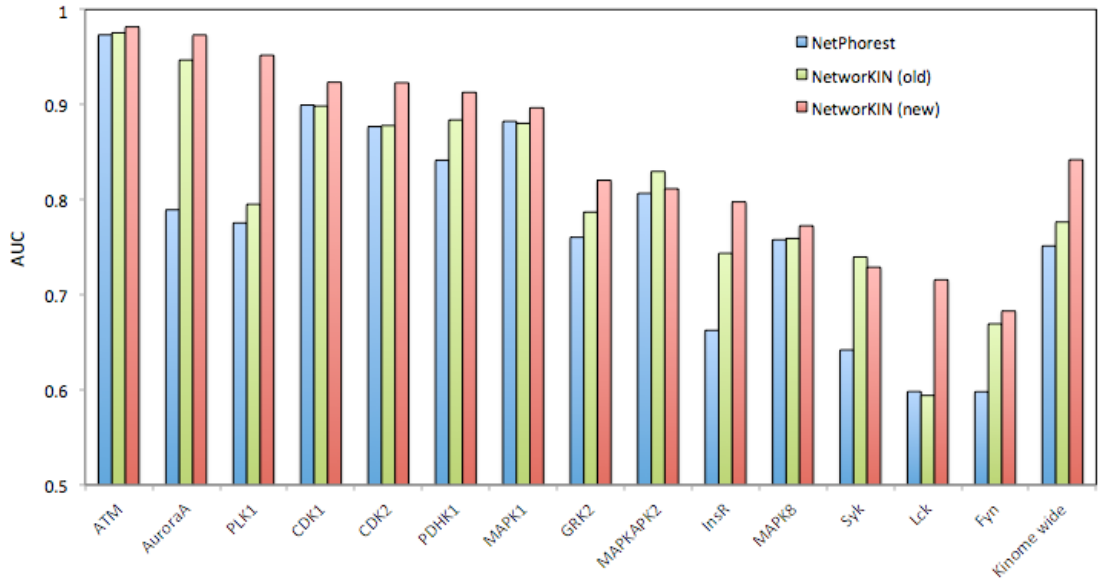
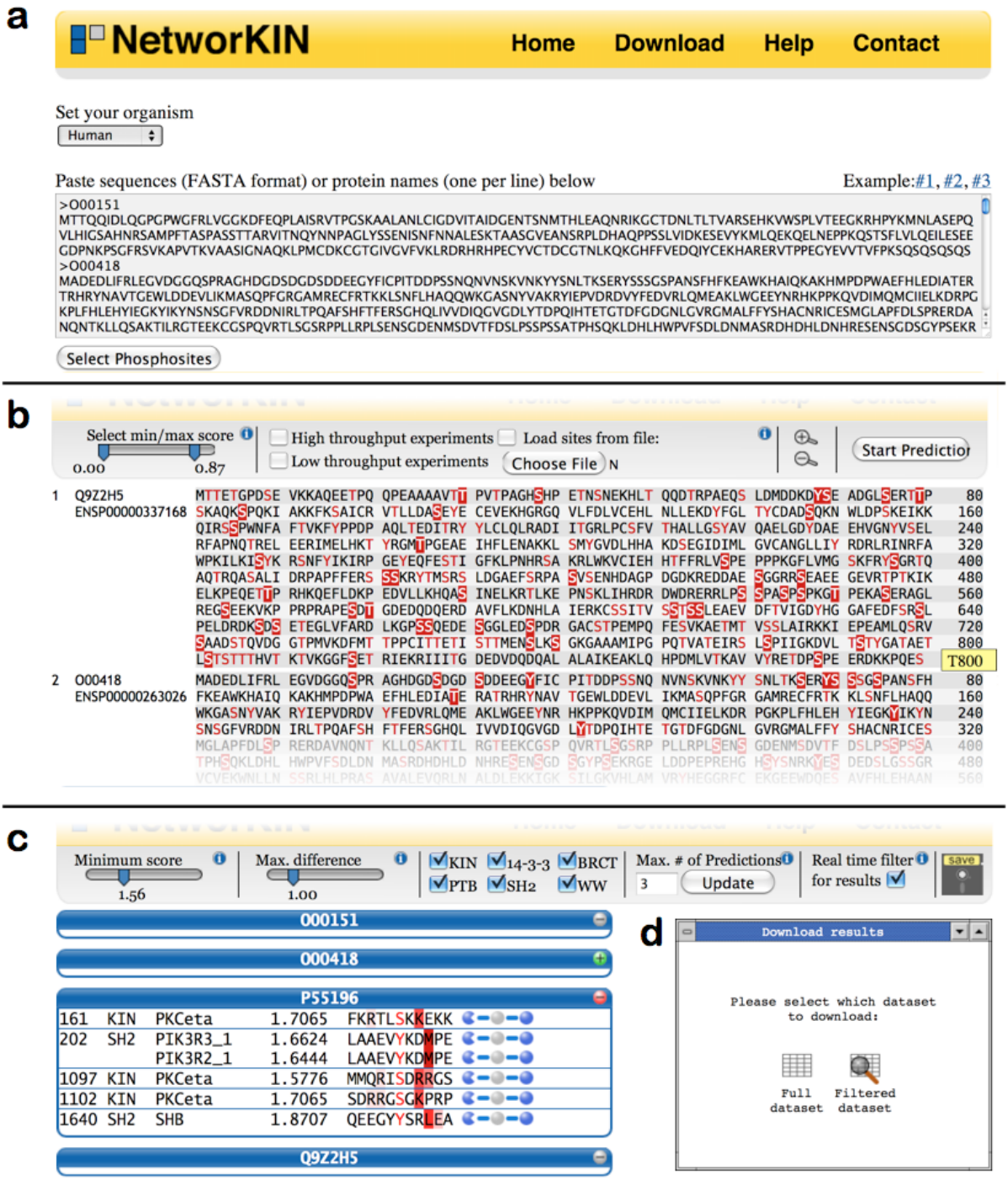


Figure 1



Supplementary Figure 1



Supplementary Figure 2

References

1. Seet, B. T., Dikic, I., Zhou, M. M. & Pawson, T. Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* **7**, 473-483 (2006).
2. Lim, W. A. & Pawson, T. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* **142**, 661-667 (2010).
3. Davis, M. I. et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1046-1051 (2011).
4. Anastassiadis, T., Deacon, S. W., Devarajan, K., Ma, H. & Peterson, J. R. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1039-1045 (2011).
5. Dinkel, H. et al. Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic Acids Res* **39**, D261-7 (2011).
6. Linding, R. et al. Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415-1426 (2007).
7. Miller, M. L. et al. Linear motif atlas for phosphorylation-dependent signaling. *Sci Signal* **1**, ra2 (2008).
8. Bakal, C. et al. Phosphorylation networks regulating JNK activity in diverse genetic backgrounds. *Science* **322**, 453-456 (2008).
9. Tan, C. S. et al. Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci Signal* **2**, ra39 (2009).
10. Van Hoof, D. et al. Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell* **5**, 214-226 (2009).
11. Jorgensen, C. et al. Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells. *Science* **326**, 1502-1509 (2009).
12. Lundby, A. et al. In vivo phosphoproteomics analysis reveals the cardiac targets of beta-adrenergic receptor signaling. *Sci Signal* **6**, rs11 (2013).
13. Hekmat, O. et al. TIMP-1 increases expression and phosphorylation of proteins associated with drug resistance in breast cancer cells. *J Proteome Res* **12**, 4136-4151 (2013).
14. Lundby, A. et al. In vivo phosphoproteomics analysis reveals the cardiac targets of beta-adrenergic receptor signaling. *Sci Signal* **6**, rs11 (2013).
15. Olsen, J. V. et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* **3**, ra3 (2010).
16. Zanivan, S. et al. In vivo SILAC-based proteomics reveals phosphoproteome changes during mouse skin carcinogenesis. *Cell Rep* **3**, 552-566 (2013).
17. Zhou, H. et al. Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. *Mol Cell Proteomics* **10**, M110.006452 (2011).
18. Rosenqvist, H., Ye, J. & Jensen, O. N. Analytical strategies in mass spectrometry-based phosphoproteomics. *Methods Mol Biol* **753**, 183-213 (2011).
19. Franceschini, A. et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* **41**, D808-15 (2013).
20. Janes, K. A. et al. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* **310**, 1646-1653 (2005).

21. Yaffe, M. B. The scientific drunk and the lamppost: massive sequencing efforts in cancer discovery and treatment. *Sci Signal* **6**, pe13 (2013).
22. Margolin, A. A. et al. Reverse engineering cellular networks. *Nat Protoc* **1**, 662-671 (2006).
23. Tuncbag, N., McCallum, S., Huang, S. S. & Fraenkel, E. SteinerNet: a web server for integrating 'omic' data to discover hidden components of response pathways. *Nucleic Acids Res* **40**, W505-9 (2012).
24. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431-432 (2011).
25. Hornbeck, P. V. et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* **40**, D261-70 (2012).
26. Mok, J. et al. Deciphering protein kinase specificity through large-scale analysis of yeast phosphorylation site motifs. *Sci Signal* **3**, ra12 (2010).
27. Zanivan, S. et al. In vivo SILAC-based proteomics reveals phosphoproteome changes during mouse skin carcinogenesis. *Cell Rep* **3**, 552-566 (2013).
28. Pawson, T. & Linding, R. Network medicine. *FEBS Lett* **582**, 1266-1270 (2008).
29. Fedorov, O., Muller, S. & Knapp, S. The (un)targeted cancer kinome. *Nat Chem Biol* **6**, 166-169 (2010).
30. Gough, N. R. Focus issue: From genomic mutations to oncogenic pathways. *Sci Signal* **6**, eg3 (2013).
31. Obata, T. et al. Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *J Biol Chem* **275**, 36108-36115 (2000).
32. Hutti, J. E. et al. A rapid method for determining protein kinase phosphorylation specificity. *Nat Methods* **1**, 27-29 (2004).
33. Liberti, S. et al. HuPho: the human phosphatase portal. *FEBS J* **280**, 379-387 (2013).
34. Pafilis, E. et al. Reflect: augmented browsing for the life scientist. *Nat Biotechnol* **27(6)**, 508-510 (2009).
35. Sadowski, I. et al. The PhosphoGRID Saccharomyces cerevisiae protein phosphorylation site database: version 2.0 update. *Database (Oxford)* **2013**, bat026 (2013).
36. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* **294**, 1351-1362 (1999).

Supplementary text

Dataset

NetworKIN integrates two types of data; protein networks from STRING, which combines various types of data based on quality scores in a way that higher quality data has larger weight. The other one consists of kinase-substrate and phospho-binding domain interactions which have been identified in cells and have been reported in public databases (e.g. PhosphoELM, PhosphositePlus and PhosphoGRID) and position specific scoring matrices (PSSMs) which were generated by peptide array experiments in collaboration with the Yaffe[31] and Turk[32] labs. The reported kinase-substrate relationships have been manually curated from literature mining of peer-reviewed reports of individual kinases and their phosphorylation of specific downstream proteins (part of the Phospho.ELM effort), where the specificity of the kinase-substrate relationships have been carefully tested and validated. In vivo phosphatase interactions were obtained from HuPho[33]. The quality of the PSSMs was assessed based on the performance of predicting experimentally observed substrates using a rigorous computational pipeline consisting of homology reduction, data partitioning, and cross-validation as previously described in detail[7], and only high quality data entries were kept and deployed in the current release. All the peptide screening data and in vivo interactions used for training NetPhorest and NetworKIN are listed in Supplementary Table 2.

Data organization for training

The NetPhorest pipeline organizes datasets based on phylogenetic trees. In the tree-based organization of the training dataset, each node has positive and negatives sets. The data organization process starts from leaf nodes, which are individual kinases. For a kinase, the positive set consists of substrates that are reported to be phosphorylated by this kinase, and the negative set consists of substrates that are reported to be phosphorylated by other kinases. Through the automated selection procedure of the NetPhorest pipeline (as previously described[7]), positive and negative sets are assembled in a way that they are best distinctive. NetworKIN uses the same data organization to avoid overtraining. Peptide binding data is used only in NetPhorest pipeline.

Utilizing STRING context information in KinomeXplorer

In the KinomeXplorer framework, phosphorylation sites are firstly classified according to the binding motifs of the kinases in the NetPhorest atlas[7], after which these classifications are refined by integrating data about the direct and indirect protein-protein interactions between the predicted kinase and its potential substrates, as this relationship is a good predictor for in vivo association. As data sources of relevant information are sparse, it is preferable to take meta-information as a refinement, thus we decided to select the STRING database as the source for our approach. STRING integrates many different sources of information, including text-mining, pathway databases or co-expression into one probabilistic protein-protein association score, allowing us to accurately integrate the NetPhorest and STRING probabilities. As most kinase-substrate interactions are unknown, we extend the STRING coverage to include also indirect paths, which broadens the association information to e.g. scaffolding processes. When calculating indirect pathways, we penalize for the path lengths

as well as for the overall connectivity of intermediate hubs (highly connected nodes). Finally, the score combination and likelihood transformation normalize for the biases introduced by the contextual data (namely study bias and network topology), as the weighting of context and motif information is adjusted specifically for each kinase. Supplementary Fig. 1 demonstrates the additional predictive power gained by the integration of contextual information.

Naive Bayes method

We combine likelihood ratios which are derived from the NetPhorest probability and network proximity scores. The conversion of the NetPhorest probability and network proximity score was done in a kinase specific manner. To convert a score (either NetPhorest probability or network proximity) to a likelihood ratio, positive and negative sets were arranged in decreasing order by the score and a sliding window was applied. The likelihood ratio was calculated as the probability of observing the value in the positive set divided by the probability of observing the value in the negative set:

$$Ln = P(Wn|pos)/P(Wn|neg)$$

$$Ls = P(Ws|pos)/P(Ws|neg)$$

Wn and Ws are windows covering a certain range of NetPhorest probability and network proximity score of String[19].

The unified likelihood (Lu) of a certain NetPhorest probability and a certain network proximity score was calculated as the product of Ln and Ls by the joint probability rule:

$$Lu = Ln \times Ls$$

Parameter optimization

To determine the optimal combination of the two parameters used (penalty for path length and node connectivity), we benchmarked against the same collection of known phosphorylation sites used to benchmark NetPhorest. To avoid overtraining we 1) use the partitioning of phosphorylation sites into training, test, and validation sets organized by the NetPhorest training pipeline, 2) use for each site the motif score from the neural network that had the site in the test set, not the training set, 3) limit the information used from the STRING network[19] to only indirect evidence paths, and 4) optimize the parameters on the test set. This ensures that parameter optimization is performed based on sites that were not used to identify the sequence specificity of the kinase in question (i.e. were not used for training the neural networks), and that the STRING network scores are not based on the same evidence that is recorded in Phospho.ELM[5]. It also ensures that the validation set has neither been used for the training of NetPhorest or for parameter optimization of NetworKIN; it is hence an independent set on which the predictive performance can be accurately assessed.

Web interface

The KinomeXplorer web service provides an interactive and intuitive user interface (Supplementary Fig. 2). The user provides either a set of sequences in FASTA format or a

list of protein or gene names. This input is then mapped to the protein collection of the STRING database[19], using either BLAST or the Reflect web service[34] depending on whether the user submitted a set of sequences or a list of names. In case of ambiguity, the user is asked to manually select the correct protein. If the input proteins only contain phosphorylation-dependent signaling domains (such as kinase, SH2, or phosphatase domains), the user is given the option to skip directly to the results page and view predicted substrates or binding partners of these. Otherwise, the input proteins are assumed to be substrates, and the user is asked to select phosphorylation sites. This can be done by uploading a file with sites, manually selecting individual sites, or using KinomeXplorer-DB, an in-house database which integrates all known phosphorylation sites from the Phospho.ELM[5], PhosphositePlus[25], and PhosphoGRID[35] databases. The web interface also integrates the NetPhos phosphorylation-site predictor[36], which can be used to add sites that are likely to be modified and/or to remove selected sites that are unlikely to be phosphorylated. The latter feature can be used e.g. to filter out likely false positive phosphorylation sites from dated and less accurately curated large-scale data sets.

The KinomeXplorer web interface also provides a high-throughput section, which takes the output of phospho-proteomics experiments. Users can either input tab-delimited phosphosite information or the output of peptide search engines such as MaxQuant and ProteomeDiscoverer . When working with this workflow method, it is imperative to select the correct sequence database in the pull-down menu, that was used for assigning the phosphorylation site locations and protein identifiers to guarantee correct protein/ phosphosite determination.

The result page offers multiple options to filter the NetPhorest or NetworKIN predictions: The primary filtering step is done by an absolute and a relative score threshold. This allows the user to show only predictions that score above a certain threshold and to hide predictions that score considerably worse than the best prediction for the same site. Additionally, one can select for which domain-classes (e.g. kinases or SH2 domains) results should be displayed. The user has the option to download either the filtered or the full set of predictions for further analysis.

In case that a FASTA file is uploaded with the respective phosphorylation sites, the protein identifiers can be of any nature, as the protein sequences will be mapped to STRING to guarantee correct protein matching. This is to accommodate the use of any sequence database of choice (e.g. UniProt, Ensembl). The supplied protein identifiers and phosphorylation site locations will also be used in the final output, to allow the user to easily identify their experimentally determined phosphorylation events of interest. This applies to the general prediction section of the KinomeXplorer web interface, and the downloadable package version as described below.

Local version of NetworKIN and NetPhorest

We also provide a local version of NetworKIN and NetPhorest in the Download section of the web interface (http://networkin-beta.cbs.dtu.dk/download.shtml). The package contains a python script of NetworKIN and NetPhorest and the ANSI-C source code of NetPhorest and data files. NetworKIN takes a FASTA and phosphosite file as input, and predicts kinases, phosphatases, and other phospho-binding domains for the supplied phosphosites. The

phosphosite file can be either a tab-delimited text file containing protein IDs, positions, and amino acids of phosphorylated residues, or output of MaxQuant and ProteomeDiscoverer. Proteins are mapped to corresponding nodes in the STRING network using sequences in the input FASTA file, not using identifiers. Therefore, in case where standard protein identifiers are ambiguous, use of local package is recommended as the supplied sequences will be mapped to STRING proteins through BLAST. The predicted substrate-kinase list is provided with additional information such as names, protein identifiers, description, and intermediate nodes in the STRING network. Similarly, NetPhorest takes the FASTA and phosphosite file supplied by the user, and predicts kinases, phosphatases and other phospho-binding domains for all the supplied phosphosites. Exact details for installation and running of the package are provided in the README.txt file.

Chapter III

Part I

Uncovering Hidden Signaling Network Dynamics by Genome-Specific Proteomics

Uncovering hidden signaling network dynamics by genome-specific proteomics

Erwin M. Schoof*[1], Pau Creixell*[1], Adrian Pasculescu*[2], Agata Wesolowska-Andersen[3], Jinho Kim[1], Ramneek Gupta[3] and Rune Linding[1@]

[1]Cellular Signal Integration Group (C-SIG), Center for Biological Sequence Analysis (CBS), Department of Systems Biology, Technical University of Denmark (DTU), Building 301, DK-2800, Lyngby, Denmark. [2]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [3]Functional Human Variation Group, Center for Biological Sequence Analysis (CBS), Department of Systems Biology, Technical University of Denmark (DTU), Building 301, DK-2800, Lyngby, Denmark. *These authors contributed equally to this work. [@]Correspondence should be addressed to R.L. (linding@cbs.dtu.dk).

Network biology aims to predict phenotype from multi-scale models of cellular information processing, by integration of quantitative genome-scale data. While Mass Spectrometry enables comprehensive sampling of cellular (phospho-)proteomes, the use of wild-type reference sequences results in masking of mutation-associated signaling events. Here we present an integrative strategy combining MS with NGS exome sequencing to perform genome-specific proteomic analysis and make cancer cell-line search databases available to the community. Deploying the approach on several cancer cell lines, we uncover otherwise-hidden signaling networks spanning many kinase-substrate interactions and a direct correlation between the fraction of sequencing reads reporting a variant allele and the likelihood of a mutation to be observed by MS. Additionally, we find a significant increase in the number of mutant peptides detected by MS compared to wild-type, suggesting a potential up-regulation of mutant protein expression in cancer cells. In conclusion, we show that genome-specific proteomics experiments enable both orthogonal cross-validation of DNA mutations and monitoring of the dynamics of signaling networks specifically dysregulated through mutations.

The fields of proteomics and genomics provide complementary views that are essential to integrate in order to predict and understand cellular phenotypes. Next Generation Sequencing (NGS) can provide information about mutations occurring throughout the entire genome, and recent advances in the MS field have led to a significant portion of the expressed proteome to be readily observable[1-5]. Additionally, through optimized enrichment procedures, a large set of PTMs (such as phosphorylation, acetylation and ubiquitination) can be analyzed and quantified through MS based studies. Thus, by identification and quantitation of thousands of modified peptides in a single experiment, it is possible to globally monitor altered signaling dynamics of mutated protein entities in any given biological system. Combined, these technologies pave the way to genome-wide investigations of how mutations at the DNA level are propagated at the protein level, and how the cell may regulate their expression. This provides a starting point to make inferences about the functional effects of such lesions, thus shaping a far more complete picture of the functional impact of mutations than genomic or proteomic studies alone[6]. Before conclusions can be drawn about the expression of mutations at the protein level, and their potential role in altering cellular information processing, it is imperative they are directly observed experimentally. Here, we present a methodology for including prior knowledge about the genome of a biological system being probed when conducting mass spectrometry (MS) based proteomics studies. By opting for next-generation DNA sequencing (as opposed to RNA-seq) as our sequencing technology, we expand the number of mutations we can explore to include genes not transcribed at any specific time or condition[7]. While custom search databases have been used in the past[8-11], conducting exome-sequencing experiments in combination with global (phospho-)proteomics experiments to investigate the expression of mutations at the proteome level and the dynamic modulation of mutation-flanking phosphorylation sites in cancer cell lines has thus far not been demonstrated.

Traditionally, the raw data produced by an MS experiment, representing the peptides present within a sample, is matched to a database of reference protein sequences, in order to identify the observed peptide spectra using hypothetical spectra derived from *in silico* digestions (Figure 1A, C & E). If a mutation occurs in these proteins however, the experimental spectra will not match with their theoretical spectra, leading to either misidentification or lack of identification of these proteins. As many mutations have been attributed to play a role in disease[12-14], it is imperative that the protein dynamics associated with these mutations can be studied. Here, we demonstrate that through the use of exome-wide, genome-specific spectra databases, obtained from accompanying NGS experiments, we can improve the number of identified proteins and phosphorylation sites due to identification of the mutated proteins and peptides, paving the way for more accurate and comprehensive modeling of signaling networks (Figure 1B, D & F). Additionally, as the functional impact of a mutation can be fundamental to a given disease phenotype, monitoring the

dynamics of mutations and phosphorylation events flanking these mutations at the protein level is essential for a more thorough understanding of cellular disease mechanisms. As a proof of principle, we here deployed the HT-29, HeLa and MCF7 cancer cell lines. These commonly-used cancer cell lines have previously been broadly characterized in terms of copy number variation, mRNA expression, phospho-proteomics and morphology[4,15-21]. In the case of HeLa, even its full exome has recently been sequenced[22]. However, to date no study has tried to integrate global datasets originating from both MS and DNA sequencing on these cell lines, to investigate how mutations are propagated at the proteome level. We analyzed the cell lines through deep (phospho-)proteomic and genomic analysis. Using the Q-Exactive Orbitrap platform (Thermo Fisher Scientific) we identified, on average, 8,012 unique proteins and 10,008 unique Class I phosphorylation sites in the different cell lines (see Table 1 for details). Using the HiSeq platform (Illumina), we performed exome sequencing with an average depth of 80X and >95% of all reads at 10X or more. This resulted, on average, in the identification of 9,133 missense variants, equating to 5,317 altered mutated protein sequences with respect to the human reference genome (see Table 2 for details).

The MS spectra generated by the proteomics experiments were searched using two variations of the Ensembl v68 sequence database: one containing only the wild-type reference sequences, and one containing both the wild-type and detected mutant variants of the proteins. These sequence databases were constructed separately in a cell-line specific manner, to avoid an increase in false discovery rates due to expanding the search space unnecessarily[23,24]. The increase in the number of entries of the sequence databases was on average of 9.4%. When using the reference Ensembl database, we were generally able to identify less proteins and phosphorylation sites, compared to including the mutant variants within the search database. In total, we identified an average of 220 additional proteins and 418 confidently localized additional phosphorylation sites by utilizing the genome-specific information for analyzing the proteomics data instead of the reference Ensembl database (see Table 1 for details). This is a 3-4% increase in identifications compared to using the reference database alone, and more importantly, opens the possibility for looking at the dynamics of these mutant proteins and phosphorylation sites, which would have otherwise been rendered undetectable due to sequence variation. Additionally, the proteomics approach enables one to distinguish technical artifacts from real variants present in the genome of the biological samples being studied.

Next, we attempted to investigate a long-standing question of how mutations are propagated at the proteome level. We analyzed whether the variant allele frequency (VAF) originating from NGS data, could be correlated to the likelihood of observing the mutation in the proteomics data. From the distribution of peptides (Figure 2A) containing a mutation with a given fraction of reads reporting a variant allele, it is clear that, in all three cell lines, this measure is directly related to the expression rate of the peptides bearing this mutation. In other words, the higher the number of NGS reads of a given mutant allele, the higher the likelihood of identifying the mutant peptide using mass spectrometry, providing orthogonal validation whether a given mutation is present at the genomic level and also expressed at the proteome level.

In order to investigate how many of the variant sites were actually expressed in the cells, we generated all possible tryptic peptides *in silico*, in order to see what percentage was observed. As can be seen in Figure 2B, while exact percentages vary slightly between the cell lines, overall we only observe an average of 3% of all possibly observable mutant peptides. Compared to the peptide coverage of the non-mutated proteome (1% on average), this percentage is significantly higher than expected by chance. This may suggest that mutant proteins show a higher degree of expression than wild-type proteins, rendering them more detectable in the MS experiment. Validation experiments in this regard are currently on-going, as this argues against what has previously been published in the literature; that mutant proteins show a lower degree of expression than wild-type proteins[25].

By combining the two technologies as described, our method provides a platform to conduct orthogonal cross-validation of mutations using NGS and MS data. More specifically, we investigated whether we could identify cases where there was disagreement between NGS and MS results. Focusing our attention on mutations with the highest NGS evidence of a homozygous mutation (mutations having a VAF of 1), we identified a few cases for which MS only identified the wild-type peptide (HeLa:6, MCF7:3, HT29:4). While we cannot conclude that this mutation is not present (the mutant peptide may have simply not been detected by the MS), we can conclude that these positions were incorrectly identified as homozygous mutations by the NGS data. Out of all the possible mutated peptides with a reported mutant allele frequency of 1, we detected 167 peptides in HeLa, 143 peptides in MCF7 and 109 peptides in HT29 in our MS data (Online Supplementary Table 1). As six, three and four of these peptides were found respectively only as wild-type variants, our data suggests an NGS error rate of incorrectly assigning mutations as homozygous at around 3%. Of the total number of identified proteins, an average of 261 (309 in HeLa, 219 in HT29 and 255 in MCF7)

were identified based on their mutated peptides (Online Supplementary Table 2). According to our sequencing results, the number of proteins containing at least one missense mutation, and therefore potential number of additional protein variants identifiable by MS, is on average 5,317 for the cell lines investigated in this study. It is well established that a single amino acid variant can have a significant impact on protein function (e.g.[26,27]), hence underlining the importance of being able to observe these mutant variants and study their dynamics in the cell at the protein level.

Additionally, we sought to investigate mutations hitting the region surrounding observed phosphorylation sites. These sites are of particular interest to labs trying to model kinase-substrate interactions, as they would be rendered unobservable using the conventional MS approach, excluding them from any subsequent modeling analysis. As the presence of these sites confirms the fact that kinases do interact with them, this underlines the importance of being able to monitor their dynamics experimentally. In total, we identified 86 mutated peptides with confidently localized phosphorylation sites using our genome-specific database (Online Supplementary Table 3). In order to assess the 'systems effect' of identifying these genome-specific phosphorylation events, we reconstructed the signaling network models containing all cell line specific phosphorylation sites that would have been missed had we not used a genome-specific proteomics approach. By computational modeling of the upstream kinases using NetworKIN[28,29] (Figure 3), it seems that, in all three cell lines, several PKC-family members interact with a number of proteins harboring a mutation, indicating a potential involvement in transcriptional regulation, cell migration, and drug resistance[30-33]. Additionally, several cell cycle related kinases such as MAPK1, MAPK3 and CDK-family kinases seem to interact with a subset of mutated proteins, suggesting these mutations may affect cell cycle or mitosis related signaling in these cell lines. Furthermore, Casein Kinase 1 also interacts with several mutated proteins, which may decrease the cell-cell-adhesion dependence and TRAIL-induced apoptosis of these cell lines[34,35]. Finally, the modulation of mutation-specific phosphorylation events by PAK1 may underlie the pro-survival phenotype associated with this kinase[36-39].

In this study, we have provided additional evidence for the importance of integrating different types of 'omics data, in order to obtain an accurate foundation for the reconstruction of cellular signaling networks. Additionally, we have made all the FASTA files available for download (http://www.lindinglab.org/GSP/index.html), so that genome-specific searches can be conducted on commonly used cancer cell lines by the community. This resource will be kept up-to-date with newly sequenced cell lines as data is made available. Due to recent advances in MS technology, it is now possible to obtain deep coverage of the proteome and phospho-proteome, which can be complemented by exome-wide deep sequencing data. We have here generalized and extended the concept of taking into account systems-wide genome-specific protein sequence information when conducting MS experiments, allowing the identity and dynamics of the mutant proteome to be investigated. In order to assess the functional impact of mutations at a systems level, MS is a key technology, as it allows the analysis of tens of thousands of proteins and phosphorylation sites from a single sample. Considering the ever improving dynamic range in mass spectrometry, the number of observed mutant peptides, while currently relatively modest though significant, will increase in future studies. We demonstrate that conducting genome-specific proteomics experiments is now feasible, even in an un-targeted, global MS setting. It is likely that additional benefits could be gained by deploying a targeted MS approach. Targeted proteomics such as SRM[40,41] may be the best proteomics strategy to monitor mutant peptides and proteins, as global approaches can currently not guarantee that this specific part of the proteome will be represented in the MS results due to the inherent dynamic range limitation. This also likely explains why a large proportion of the mutations reported by the sequencing data could not be observed in the global MS results.

Based on our comparison of experimentally observed wild-type versus mutant peptides, where we conclude that only ~3% of all mutant peptides are currently observable, the total number of possibly observable mutant peptides can be up to 30-fold higher than reported in this study. Given the rapid progression in MS and NGS technology, the need for, and benefit of this method is likely to increase significantly in future personalized network medicine studies[14,42,43], where patient samples can undergo NGS and MS experiments to study a disease from the genomic and proteomic perspective, in order to guide the best possible therapeutic strategies. Additionally, it will most likely prove useful in distinguishing between key mutations driving a given disease state, or mutations arising sporadically. In conclusion, through the method described here, one can use the knowledge gained from NGS experiments in order to improve the sensitivity and accuracy of MS experiments, rendering the two technologies a very powerful combination for investigating complex diseases such as cancer, diabetes or neurological illnesses.

FIGURE LEGENDS

FIG.1

A & B) Limitations of unspecific MS

A) Conceptual overview of how a mutated protein is identified as a wild-type protein in an unspecific database search. Due to the lack of the mutated peptide in the reference database, only wild-type peptides are used for matching to the parent protein. B) Only when a genome-specific database is used, can the mutated peptide be matched to its parent sequence and is the correct variant of the protein identified.

C & D) Example of genome-specific mutant peptide identified with our approach.

MS spectra of wild-type (E) and mutant (F) versions of the same peptide. The mutant peptide becomes identifiable due to using an HT-29-specific database for conducting the MS data search.

E & F) Unspecific versus Genome-specific MS approach.

As opposed to previous unspecific MS approaches (C), our genome-specific approach (D) allows for a sample-specific search of MS data by exome sequencing the sample and generating a specific database. This approach allows the identification of mutant proteins that would otherwise be hidden and avoids the mismatching of spectra caused by the absence of a given mutant gene in standard reference databases.

FIG.2

A) Comparison of reads reporting a mutant allele and MS observability.

As can be observed in this graph, the higher the fraction of reads that report a mutation, the higher the likelihood that this mutant peptide is observed in the MS data.

B) Mutant vs Wild-Type peptide coverage by MS

Boxplots showing fraction of mutated and wild-type peptides that were observed by mass spectrometry.

FIG. 3

Hidden (phospho-) proteome

Phosphorylation-based signaling networks that became apparent only when using the genome-specific approach. For all the mutation-flanking phosphorylation sites, modulating kinases were predicted using the KinomeXplorer framework, and the mutant phosphorylated proteins and upstream kinases are represented in red and blue respectively.

ACKNOWLEDGMENTS

We would like to thank members of the Linding Lab and the Erler Lab (BRIC, Denmark) for useful input on the manuscript. This work was supported by the Lundbeck Foundation and the Human Frontier Science Program.

AUTHOR CONTRIBUTIONS

R.L. conceived the project. E.M.S., P.C., A.P. and R.L. designed the experiments. E.M.S., P.C. and A.P. performed the experiments. E.M.S., P.C., A.P., A.W.A and J.K. analyzed the data, and E.M.S., P.C., A.P. and R.L. wrote the paper. R.G. supervised the genomic analysis. R.L. oversaw the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Beck, M. et al. The quantitative proteome of a human cell line. *Mol Syst Biol* **7**, 549 (2011).
2. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics* **11**, M111.014050 (2012).
3. Munoz, J. & Heck, A. J. Quantitative proteome and phosphoproteome analysis of human pluripotent stem cells. *Methods Mol Biol* **767**, 297-312 (2011).
4. Nagaraj, N. et al. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* **7**, 548 (2011).
5. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).
6. Krug, K., Nahnsen, S. & Macek, B. Mass spectrometry at the interface of proteomics and genomics. *Mol Biosyst* **7**, 284-291 (2011).
7. Ku, C. S. et al. Exome versus transcriptome sequencing in identifying coding region variants. *Expert Rev Mol Diagn* **12**, 241-251 (2012).
8. Branca, R. M. et al. HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat Methods* **11**, 59-62 (2014).
9. Cheung, W. C. et al. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat Biotechnol* **30**, 447-452 (2012).
10. Low, T. Y. et al. Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. *Cell Rep* **5**, 1469-1478 (2013).
11. Sheynkman, G. M., Shortreed, M. R., Frey, B. L., Scalf, M. & Smith, L. M. Large-scale mass spectrometric detection of variant peptides resulting from nonsynonymous nucleotide differences. *J Proteome Res* **13**, 228-240 (2014).
12. Greenman, C. et al. Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-158 (2007).
13. Wong, K. M., Hudson, T. J. & McPherson, J. D. Unraveling the genetics of cancer: genome sequencing and beyond. *Annu Rev Genomics Hum Genet* **12**, 407-430 (2011).
14. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546-1558 (2013).
15. Imami, K., Sugiyama, N., Tomita, M. & Ishihama, Y. Quantitative proteome and phosphoproteome analyses of cultured cells based on SILAC labeling without requirement of serum dialysis. *Mol Biosyst* **6**, 594-602 (2010).
16. Kim, J. E., Tannenbaum, S. R. & White, F. M. Global phosphoproteome of HT-29 human colon adenocarcinoma cells. *J Proteome Res* **4**, 1339-1346 (2005).
17. Le Bivic, A., Hirn, M. & Reggio, H. HT-29 cells are an in vitro model for the generation of cell polarity in epithelia during embryonic differentiation. *Proc Natl Acad Sci U S A* **85**, 136-140 (1988).
18. Petretti, T., Kemmner, W., Schulze, B. & Schlag, P. M. Altered mRNA expression of glycosyltransferases in human colorectal carcinomas and liver metastases. *Gut* **46**, 359-366 (2000).
19. Reichelt, W. H., Liu, Y., Luna, L., Eigjo, K. & Reichelt, K. L. Early oncogene mRNA expression in HT-29 cells treated with the endogenous colon mitosis inhibitor pyroglutamyl-histidyl-glycine. *Anticancer Res* **22**, 991-996 (2002).
20. Shadeo, A. & Lam, W. L. Comprehensive copy number profiles of breast cancer cell model genomes. *Breast Cancer Res* **8**, R9 (2006).
21. Yasui, K. et al. Alteration in copy numbers of genes as a mechanism for acquired drug resistance. *Cancer Res* **64**, 1403-1410 (2004).
22. Landry, J. J. et al. The genomic and transcriptomic landscape of a HeLa cell line. *G3 (Bethesda)* **3**, 1213-1224 (2013).

23. Bunger, M. K. et al. Detection and validation of non-synonymous coding SNPs from orthogonal analysis of shotgun proteomics data. *J Proteome Res* **6**, 2331-2340 (2007).
24. Li, J. et al. A bioinformatics workflow for variant peptide detection in shotgun proteomics. *Mol Cell Proteomics* **10**, M110.006536 (2011).
25. Shah, S. P. et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395-399 (2012).
26. Davies, H. et al. Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-954 (2002).
27. Songyang, Z. et al. Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. *Nature* **373**, 536-539 (1995).
28. Linding, R. et al. Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415-1426 (2007).
29. Linding, R. et al. NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res* **36**, D695-9 (2008).
30. Carey, I., Williams, C. L., Ways, D. K. & Noti, J. D. Overexpression of protein kinase C-alpha in MCF-7 breast cancer cells results in differential regulation and expression of alphavbeta3 and alphavbeta5. *Int J Oncol* **15**, 127-136 (1999).
31. Johnson, C. L., Lu, D., Huang, J. & Basu, A. Regulation of p53 stabilization by DNA damage and protein kinase C. *Mol Cancer Ther* **1**, 861-867 (2002).
32. Lee, S. A., Karaszkiewicz, J. W. & Anderson, W. B. Elevated level of nuclear protein kinase C in multidrug-resistant MCF-7 human breast carcinoma cells. *Cancer Res* **52**, 3750-3759 (1992).
33. Masur, K., Lang, K., Niggemann, B., Zanker, K. S. & Entschladen, F. High PKC alpha and low E-cadherin expression contribute to high migratory activity of colon carcinoma cells. *Mol Biol Cell* **12**, 1973-1982 (2001).
34. Dupre-Crochet, S. et al. Casein kinase 1 is a novel negative regulator of E-cadherin-based cell-cell contacts. *Mol Cell Biol* **27**, 3804-3816 (2007).
35. Izeradjene, K., Douglas, L., Delaney, A. B. & Houghton, J. A. Casein kinase I attenuates tumor necrosis factor-related apoptosis-inducing ligand-induced apoptosis by regulating the recruitment of fas-associated death domain and procaspase-8 to the death-inducing signaling complex. *Cancer Res* **64**, 8036-8044 (2004).
36. Coniglio, S. J., Zavarella, S. & Symons, M. H. Pak1 and Pak2 mediate tumor cell invasion through distinct signaling mechanisms. *Mol Cell Biol* **28**, 4162-4172 (2008).
37. Li, Q., Mullins, S. R., Sloane, B. F. & Mattingly, R. R. p21-Activated kinase 1 coordinates aberrant cell survival and pericellular proteolysis in a three-dimensional culture model for premalignant progression of human breast cancer. *Neoplasia* **10**, 314-329 (2008).
38. Porcu, G. et al. Combined p21-activated kinase and farnesyltransferase inhibitor treatment exhibits enhanced anti-proliferative activity on melanoma, colon and lung cancer cell lines. *Mol Cancer* **12**, 88 (2013).
39. Sun, J., Khalid, S., Rozakis-Adcock, M., Fantus, I. G. & Jin, T. P-21-activated protein kinase-1 functions as a linker between insulin and Wnt signaling pathways in the intestine. *Oncogene* **28**, 3132-3144 (2009).
40. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **138**, 795-806 (2009).
41. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci U S A* **104**, 5860-5865 (2007).
42. Creixell, P., Schoof, E. M., Erler, J. T. & Linding, R. Navigating cancer network attractors for tumor-specific therapy. *Nat Biotechnol* **30**, 842-848 (2012).
43. Pawson, T. & Linding, R. Network medicine. *FEBS Lett* **582**, 1266-1270 (2008).

ONLINE METHODS

Sample preparation for sequencing and data analysis. HT-29, HeLa and MCF7 cells were obtained from ATCC and grown to 80% confluency in a T-75 flask, and DNA extraction was performed using reagents and instructions provided with the Qiagen QIAamp DNA Mini kit. 5 ug of purified DNA were sent to Roche Nimblegen for full exome sequencing using the SeqCap EZ Human Exome Library v3.0 capture kit. High-quality reads, with > 80x mean coverage and > 95% of exome bases at 10x coverage, were obtained from sequencing and aligned to the NCBI37 reference human genome (version GRCh37) using the Burrows–Wheeler Alignment Tool. The alignment was refined by means of quality score recalibration and around indel realignment using Genome Analysis ToolKit package. SNP calling was performed with SAMtools package using default settings. Next, results were further filtered with VCFtools using standard default settings as well as a minimum 10x sequencing depth threshold set for SNP calling. The data was further analyzed with the help of SAMtools and BEDtools packages and custom-written Perl and Python scripts. Finally, fasta files for both wild-type and mutant protein sequences were generated using the Variant Effector Predictor (VEP) package from Ensembl.

Sample preparation for (phospho-)proteomics. HT-29, HeLa and MCF7 cells (obtained from ATCC and regularly checked for mycoplasma contamination) were grown to ~80% confluency in 15cm dishes to provide enough starting material for the phospho peptide enrichment in duplicate (24mg per repeat). Cells were lysed with ice-cold modified RIPA buffer supplemented with Roche complete protease inhibitor cocktail tablets and ß-glycerophosphate (5mM), NaF (5mM), Na-orthovanadate (1mM, activated). Lysates were sonicated on ice and spun down at 4,400xg for 20mins at 4°C. Proteins were precipitated over-night in ice cold Acetone at -20°C, and dissolved in 6M Urea, 2M Thiourea, 10mM HEPES pH 8.0. Proteins were reduced with 1mM DTT for 1hr, and alkylated with 5.5mM Chloroacetamide for 1hr, after which they were pre-digested with Lysyl Endopeptidase (Wako) at a 1:200 enzyme-to-protein ratio for 4hrs at room temperature (RT). Lysates were diluted 1:4 with 50mM Ammonium Bicarbonate, after which Trypsin (MS grade, Sigma) was added at a 1:200 enzyme-to-protein ratio and left rotating over-night at RT. Enzymatic activity was quenched by adding TFA to a final concentration of 2%, after which the samples were clarified by spinning down at 2,000xg for 5 minutes and desalted using 360mg SepPak columns (Waters WAT020515). Peptides were eluted using 2x 2mL of 40% AcN, 0.1% TFA, and 1x 2ml of 60% Acetonitrile, 0.1% TFA. For the global, Titanium Dioxide (TiO2) based phospho peptide enrichment, the eluent was directly subjected to SCX fractionation, where peptides were separated over a 0-30% Buffer B gradient in 60 minutes at a 1ml/min flowrate (Buffer A: 5mM potassium dihydrogen phosphate, 30% Acetonitrile, pH2.7; Buffer B: 5mM potassium dihydrogen phosphate, 30% Acetonitrile, 350mM potassium chloride, pH2.7). The resulting fractions were pooled according to their chromatography into 11 final samples, which were enriched seperately for phosphorylated peptides. Six aliquots were taken at this point for the global proteome analysis. The TiO2 enrichment was conducted similarly to [Olsen et al., MSPP 2009], with several adjustments. For the TiO2 loading solution, 0.02g/ml dihydrobenzoic acid was dissolved in 30% Acetonitrile and 4% TFA, and the TiO2 beads were incubated in this solution for 15 minutes prior to peptide enrichment. Each pooled SCX fraction was enriched with 1.7mg of TiO2 beads suspended in 6ul of TiO2 loading solution, and left to rotate end-over-end for 30 minutes at RT. The flow-through (early eluting fractions) was enriched three times consecutively, whereas the single SCX chromatography peak peptide samples were enriched twice. Samples were spun at 2000xg for 5 minutes (RT), and pelleted beads were washed with 100ul SCX Buffer B. Subsequently, beads were pelleted again (2000xg, 5minutes, RT) and washed with 100ul 40% Acetonitrile, 0.25% acetic acid, 0.5% TFA. Finally, pelleted beads were re-suspended in 50ul 80% Acetonitrile, 0.5% acetic acid, and transferred to separate in-house packed C8 StageTips [Rappsilber et al., Nat Protoc 2007]. Liquid was spun through at 3000 rpm for 1 minute, after which the phosphorylated peptides were eluted with 1x 20ul 5% Ammonia and 1x 20ul 10% Ammonia, 25% Acetonitrile into a 96-well PCR plate, containing 20ul of 1% TFA, 5% Acetonitrile solution. Peptides were concentrated to a total volume of 10ul in an Eppendorf Speedvac, and acidified with 40ul of 1% TFA, 5% Acetonitrile, after which they were desalted on in-house packed C18 StageTips prior to LC-MS analysis.

For LC-MS analysis, peptides were eluted from the StageTip with 2x 20ul 80% Acetonitrile, 0.1% Formic acid, and concentrated to 5ul final volume. The eluent was acidified with 1% TFA, 2% Acetonitrile and loaded onto a 50cm C18 EasySpray column (Thermo, ES803), using the Thermo EasyLC 1000 uHPLC system and the column oven operating at 45°C. Peptides were eluted over a 250 minute gradient, ranging from 6-60% of 80% Acetonitrile, 0.1% Formic acid, and the Q Exactive (Thermo) was run in a DD-MS2 top10 method. Full MS spectra were collected at a resolution of 70,000, with an AGC target of 3e6 or maximum injection time of 20ms and a scan range of 300-1750 m/z. The MS2 spectra were obtained at a resolution of 17,500, with an AGC target value of 1e6 or maximum injection time of 80ms.

Dynamic exclusion was set to 20s, and ions with a charge state < 2 or unknown were excluded. For the proteome samples, the settings were the same, except for a gradient time of 240mins, maximum MS2 injection time of 60ms and dynamic exclusion of 45s.

Computational analysis of MS data. In order to investigate the effect of using the HT-29-specific FASTA file as the MaxQuant (Version 1.2.7.4) search engine database, we performed the raw data searches in three ways: 1) standard Ensembl v.68 human FASTA, 2) standard Ensembl v.68 human FASTA + all possible single mutant proteins, and 3) all single possible mutant proteins only. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [Vizcaino et al., NAR 2013] with the dataset identifier PXD000267. Variable modifications were set as Methionine oxidation, Protein N-term acetylation and Serine/Threonine/Tyrosine phosphorylation, and Cysteine carbamidomethylation was set as a fixed modification. FDR rates were set to 1%, and the 'match between runs' functionality was activated.
Results from the three independent searches were stored in a MySQL database, and all further analysis was done using scripts written in-house on our "CoreFlow" platform, based on the R statistical package, MySQL and Python. All code and data will be released to the public upon request. Search results filtering was based on phosphorylation localization probability >= 0.75 and a minimum MaxQuant peptide ID score of 50, in order to only use high confidence identifications.
Peptide observability was calculated based on all possibly observable tryptic peptides originating from an *in silico* digest (minimum peptide length of 5 amino acids); the percentage of peptides observed was calculated using the following: peptides observed / total # of peptides observable x 100. For the percentage of Peptides Observed, the data size per bin of Variant Allele Frequency was between 8 and 97 with an average of 21, and the average of total peptides with possible mutation per bin was 800. We considered the data size to be sufficient for the estimation of the percentage of observed peptides. For the MS observability of the non-mutated peptides, we used sampling of the appropriate size from the set of all `in-silico` digested peptides. The sample size was equal to the size of the data set of MS observability of the mutated peptides. To test for statistical significance of the difference in MS observability between the mutant and wild-type peptides, we applied a Wilcoxon statistical test, which does not rely on the assumption of normality or independence between data sets.
The NetworKIN modeling was based using an in-house up-to-date version of NetworKIN v3.0 (part of the novel KinomeXplorer framework, and the high confidence phosphorylation sites with a surrounding mutation were analyzed. Subsequently, kinase predictions were filtered to only include predictions with a score of 2 and higher in order to reduce false positives. The results were plotted in Cytoscape (http://www.cytoscape.org) [Shannon et al., Gen. Res., 2003] for visual representation.

SUPPLEMENTARY REFERENCES:

Olsen, J et al., High accuracy mass spectrometry in large-scale analysis of protein phosphorylation. *Mass Spectrometry of Proteins and Peptides*, Volume **492**, Chapter 7 (2009)
Rappsilber et al., Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* **2** (8), 1896-906 (2007)
Shannon et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**(11):2498-504
Vizcaino JA, et al. The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res*. **41**(D1):D1063-9 (2013)

Table 1.

	HeLa		HT29		MCF7	
	Wild-Type	Mutant + WT	Wild-Type	Mutant + WT	Wild-Type	Mutant + WT
Nr of Protein IDs	8,012	8,217	7,560	7,815	7,802	8,004
Nr of Phosphorylation Site IDs	7,484	7,868	14,848	15,440	6,439	6,718
Nr of Mutated Peptide IDs	0	350	0	237	0	274

Table 2.

	HeLa	HT29	MCF7
Nr of Missense Variants	10293	8186	8922
Nr of Altered Protein Sequences	5832	4940	5181

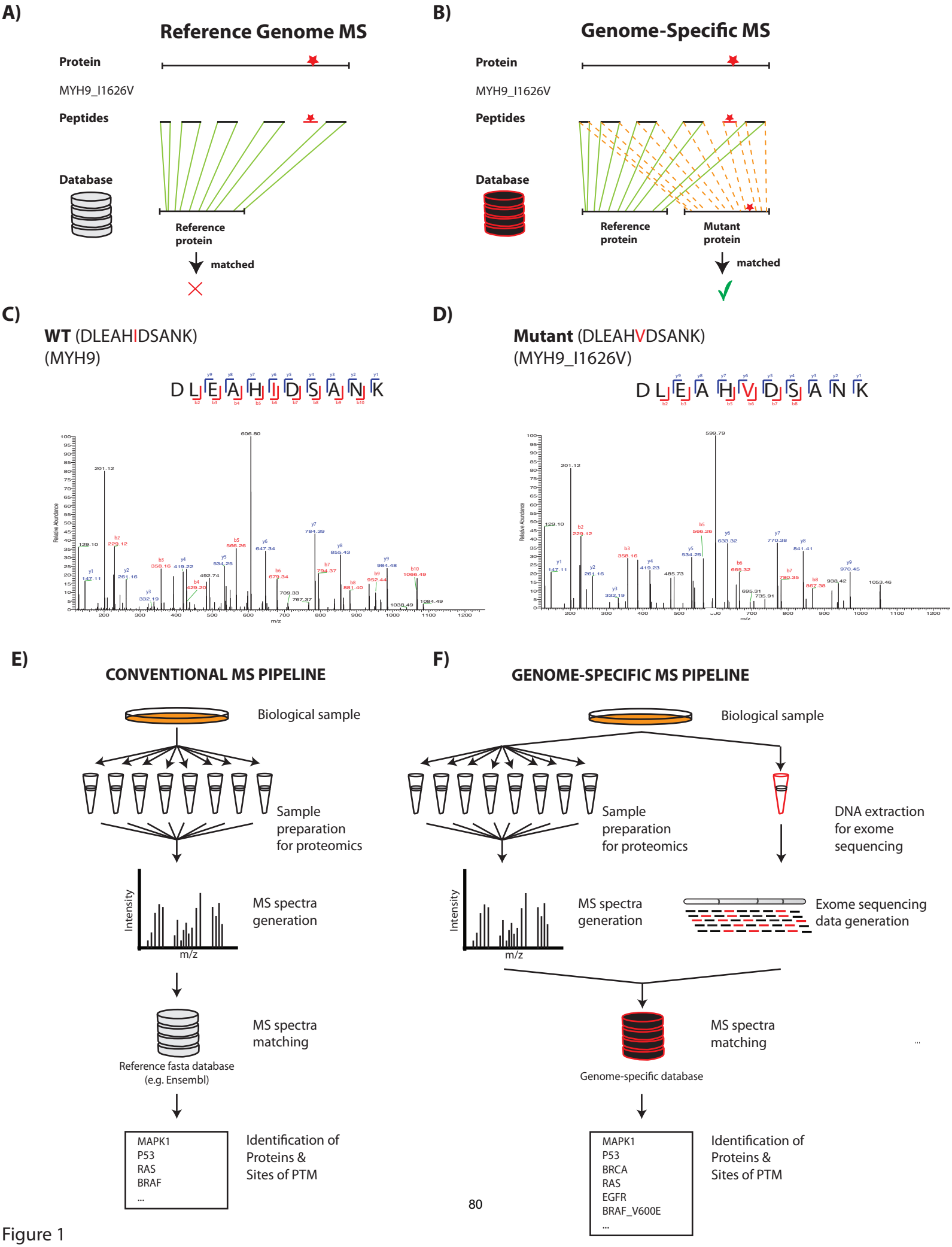
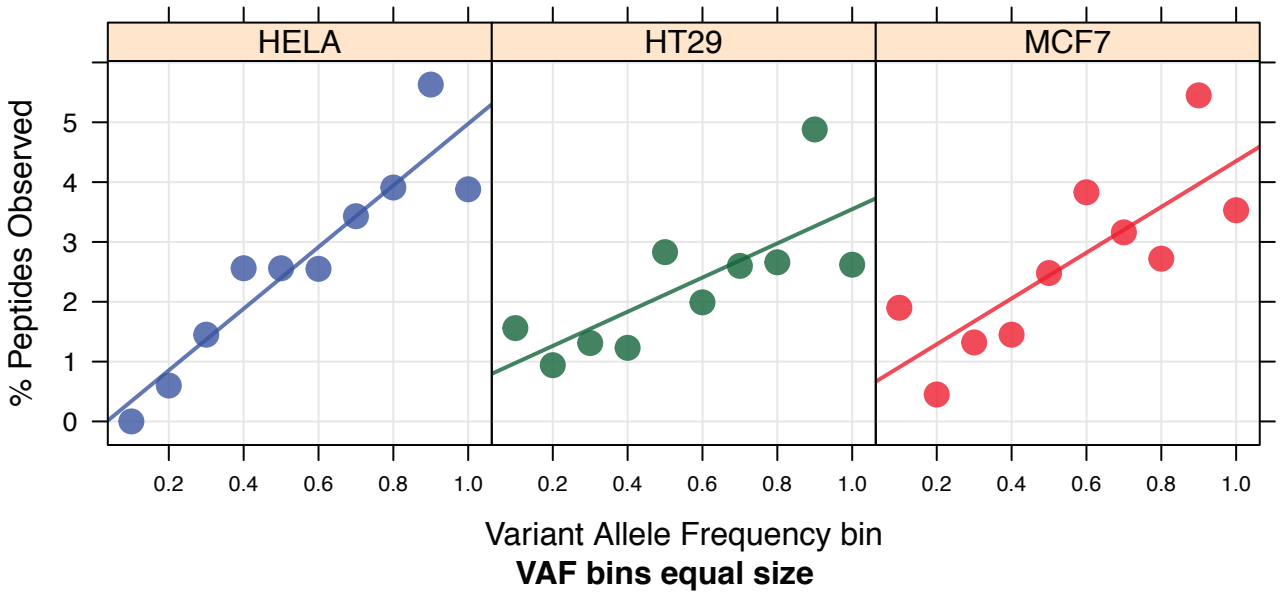

A)
Reference Genome MS
Protein
MYH9_I1626V
Peptides
Database
Reference protein
matched
B)
Genome-Specific MS
Protein
MYH9_I1626V
Peptides
Database
Reference protein
Mutant protein
matched
C)
WT (DLEAHIDSANK)
(MYH9)
D L E A H I D S A N K
D)
Mutant (DLEAHVDSANK)
(MYH9_I1626V)
D L E A H V D S A N K
E)
CONVENTIONAL MS PIPELINE
Biological sample
Sample preparation for proteomics
Intensity
m/z
MS spectra generation
MS spectra matching
Reference fasta database (e.g. Ensembl)
MAPK1
P53
RAS
BRAF
...
Identification of Proteins & Sites of PTM
F)
GENOME-SPECIFIC MS PIPELINE
Biological sample
Sample preparation for proteomics
DNA extraction for exome sequencing
Intensity
m/z
MS spectra generation
Exome sequencing data generation
MS spectra matching
Genome-specific database
MAPK1
P53
BRCA
RAS
EGFR
BRAF_V600E
...
Identification of Proteins & Sites of PTM

Figure 1

A)


HELA
HT29
MCF7
% Peptides Observed
0
1
2
3
4
5
0.2
0.4
0.6
0.8
1.0
Variant Allele Frequency bin
VAF bins equal size

B)

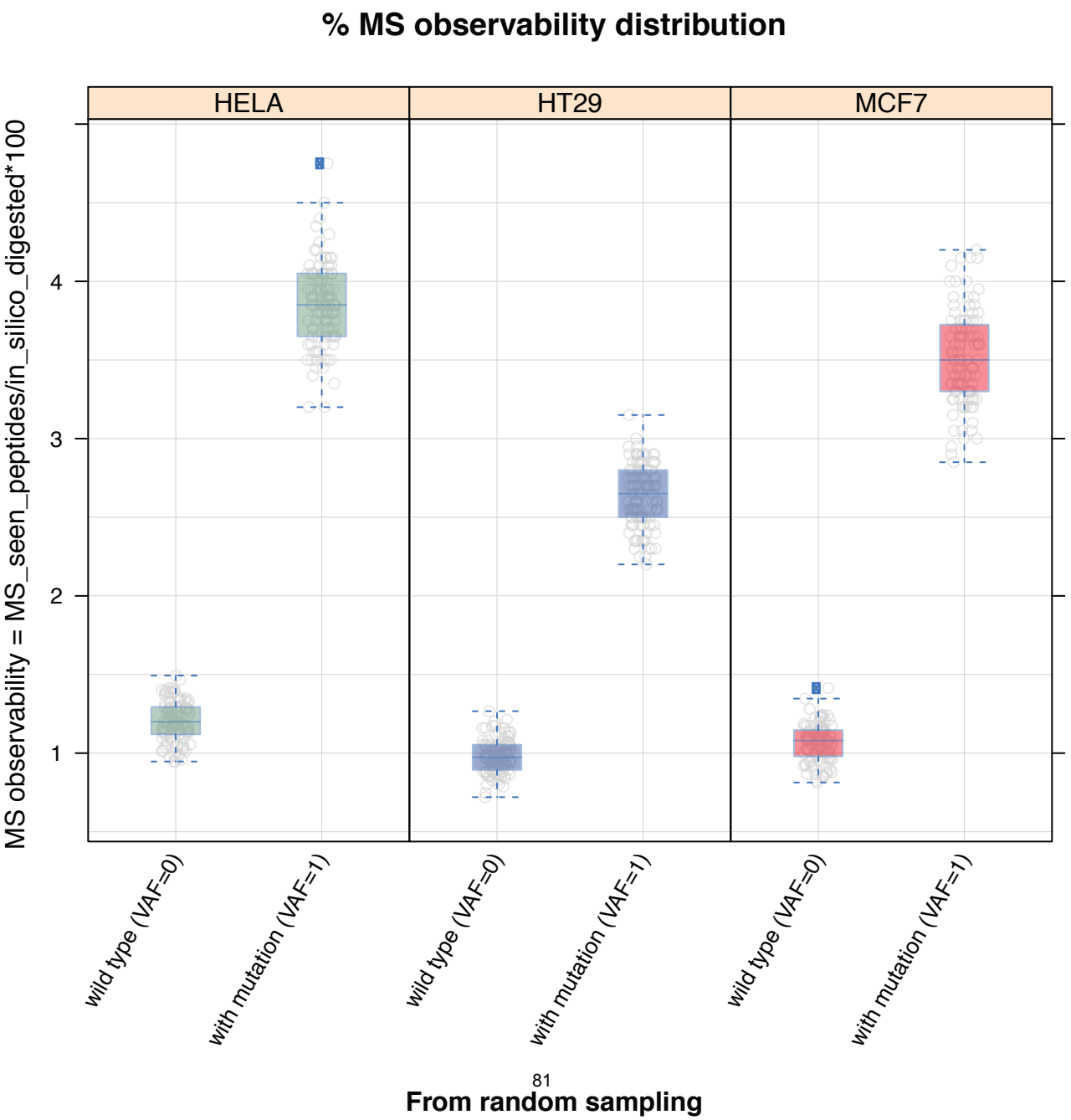
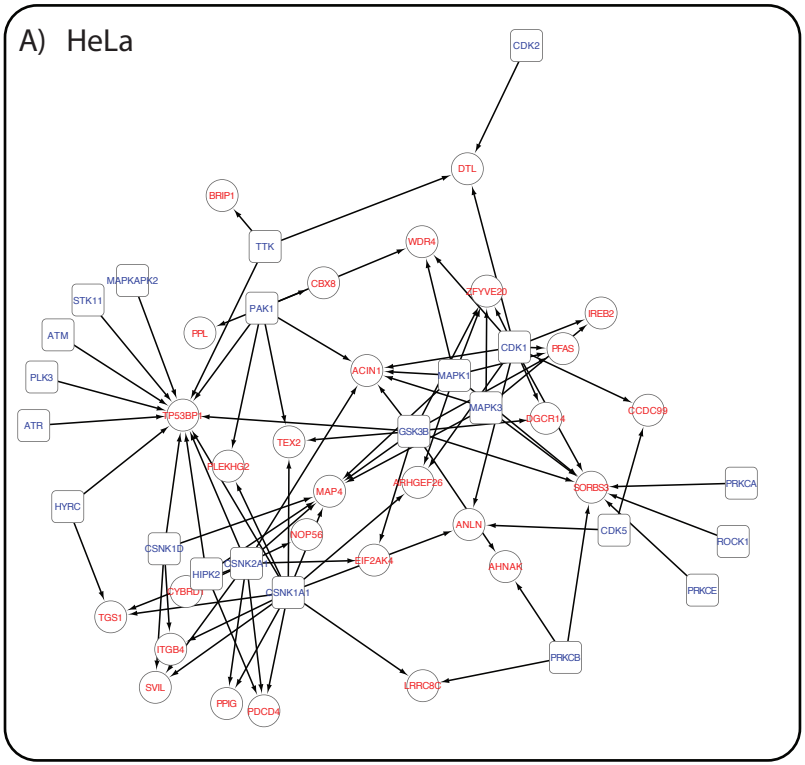
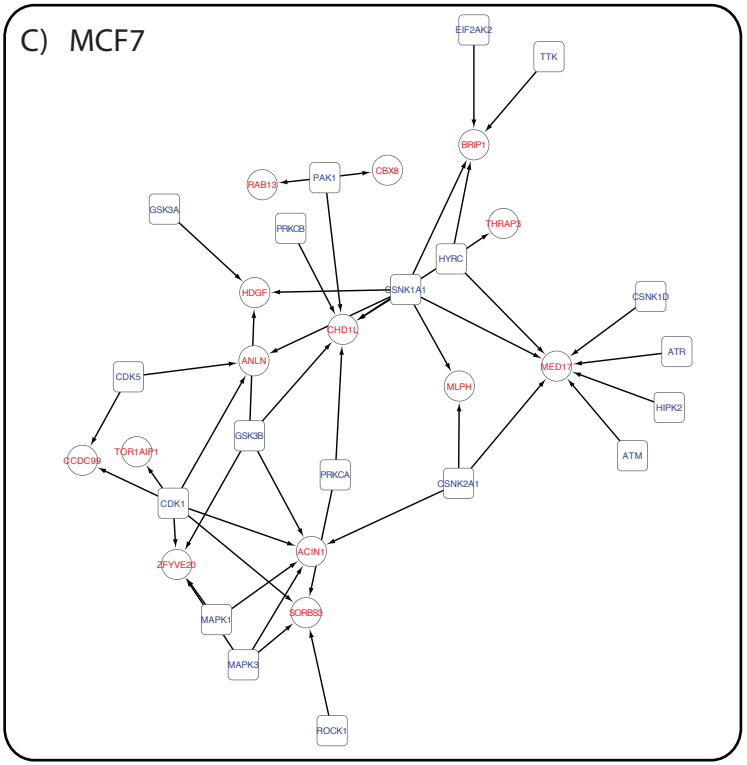

% MS observability distribution
HELA
HT29
MCF7
MS observability = MS_seen_peptides/in_silico_digested*100
1
2
3
4
wild type (VAF=0)
with mutation (VAF=1)
wild type (VAF=0)
with mutation (VAF=1)
wild type (VAF=0)
with mutation (VAF=1)
81
From random sampling

Figure 2


A) HeLa
CDK2
DTL
BRIP1
TTK
WDR4
MAPKAPK2
CBX8
STK11
PAK1
ZFYVE20
IREB2
ATM
PPL
CDK1
PFAS
ACIN1
PLK3
MAPK1
ATR
TP53BP1
MAPK3
DGCR14
CCDC99
TEX2
GSK3B
PLEKHG2
MAP4
ARHGEF26
SORBS3
PRKCA
HYRC
NOP56
ANLN
CDK5
CSNK1D
ROCK1
CSNK2A1
EIF2AK4
AHNAK
HIPK2
PRKCE
CSNK1A1
CYBRD1
TGS1
ITGB4
PRKCB
SVIL
LRRC8C
PPIG
PDCD4


C) MCF7
EIF2AK2
TTK
BRIP1
RAB13
PAK1
CBX8
GSK3A
PRKCB
THRAP3
HYRC
HDGF
SNK1A1
CSNK1D
CHD1L
ATR
ANLN
MED17
CDK5
MLPH
HIPK2
GSK3B
CCDC99
TOR1AIP1
ATM
PRKCA
CSNK2A1
CDK1
ZFYVE20
ACIN1
MAPK1
SORBS3
MAPK3
ROCK1

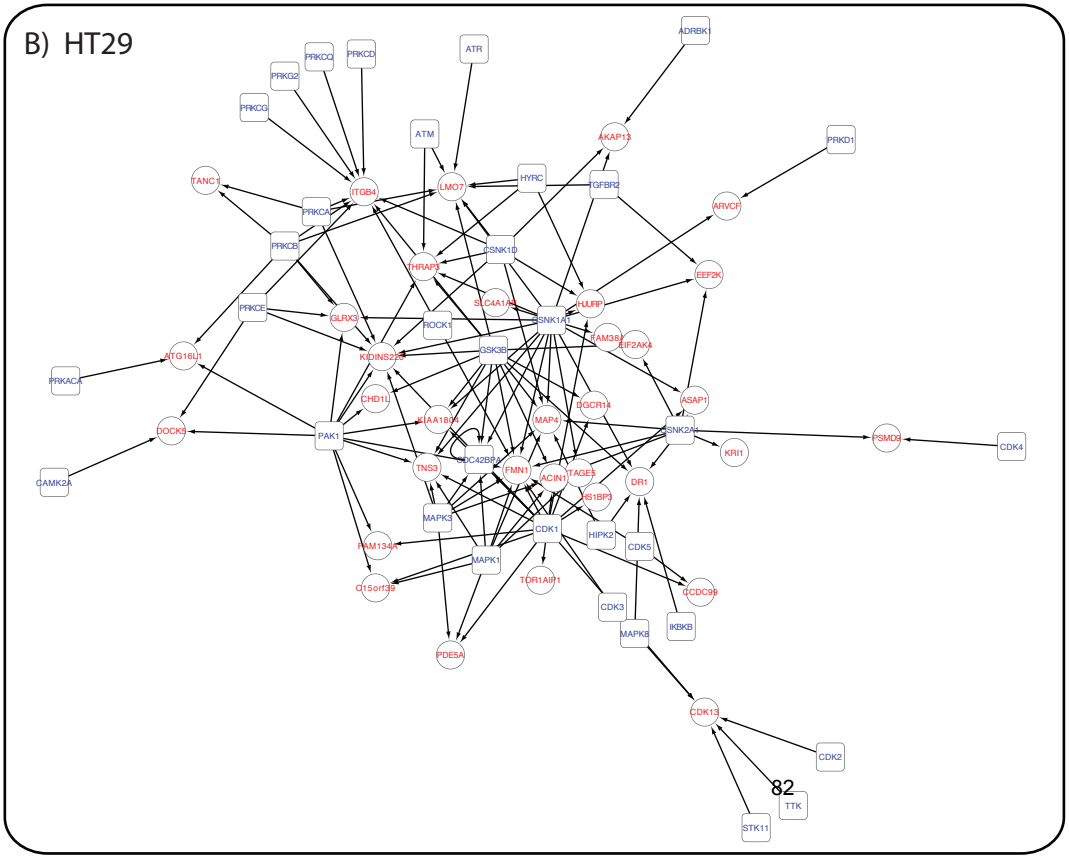

B) HT29
ADRBK1
PRKG2
PRKCQ
PRKCD
ATR
PRKCG
ATM
AKAP13
PRKD1
TANC1
ITGB4
LMO7
HYRC
TGFBR2
PRKCA
ARVCF
PRKCB
CSNK1D
THRAP3
EEF2K
SLC4A1AP
HUWE1
PRKCE
GLRX3
ROCK1
SNK1A1
TRAM38
EIF2AK4
ITG16U1
KIDINS220
GSK3B
PRKACA
CHD1L
DGCR14
ASAP1
DOCK5
KIAA1804
MAP4
SNK2A1
PSMD9
CDK4
PAK1
DDC42BP2
KRI1
TNS3
FMN1
ACIN1
TAGE3
OR1
CAMK2A
S1BP3
MAPK3
CDK1
HIPK2
CDK5
PAM134A
MAPK1
C15orf39
TOR1AIP1
CCDC99
CDK3
MAPK8
IKBKB
PDE5A
CDK13
CDK2
TTK
STK11

Figure 3

Chapter III

Part II

Modeling Colon Cancer Metastasis using Global, Quantitative and Integrative Network Biology

Modeling Colon Cancer Metastasis using Global, Quantitative and Integrative Network Biology

Erwin M. Schoof*[1], Thomas R. Cox*[2], Jesper Ferkinghoff-Borg§[1], James Longden§[1], Pau Creixell[1], Jinho Kim[1], Adrian Pasculescu[3], Cristina Costa Santini[1], Graeme I. Murray[4], Janine T. Erler$[2], Rune Linding$[1].

* these authors contributed equally

§ these authors contributed equally

$ to whom correspondence should be addressed

[1] Cellular Signal Integration Group (C-SIG), Center for Biological Sequence Analysis (CBS), Department of Systems Biology, Technical University of Denmark (DTU), Building 301, DK-2800, Lyngby, Denmark

[2] Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Denmark

[3] Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada.

4 Department of Pathology, University of Aberdeen, Scotland

Abstract:

In order to respond to alterations in its environment, a cell has to integrate multiple input-cues and modulate its signaling networks accordingly, to elicit a specific response such as proliferation or apoptosis. This process becomes significantly altered during cancer development, with genomic modifications giving rise to differential protein dynamics, ultimately resulting in disease. The exact molecular signaling networks underlying specific disease phenotypes remains elusive, as the definition thereof requires extensive analysis of not only the genomic and proteomic landscapes within a particular tumor, but also the phenotypic responses to perturbations. Here, we set out to characterize the proteomic and genomic alterations required for a metastatic phenotype, in a pair of matched cell lines, SW480 and SW620. By subsequently subjecting the cell lines to a kinome-wide RNAi screen measuring cell proliferation, we pinpoint key kinases which are involved in the survival of metastatic cells. By perturbing these kinases using Sunitinib, we are able to specifically induce apoptosis both in vitro and in vivo. The clinical relevance of this inhibitor and our cell line-based model is assessed through global assessment of the proteomic and genomic landscapes in a panel of colorectal cancer patients. In conclusion, by deploying a network biology strategy on deciphering key molecular aspects of cancer metastasis, we find some novel potential clinical targets.

Introduction:

While extensive strides have been made in recent years in terms of defining biological characteristics (termed 'Hallmarks') that drive a tumor towards malignancy[1], there is still great debate about the underlying causes. Several molecular processes are fundamental to all human disease, ranging from mutations occurring in the genome[2,3], through epigenetic regulation of DNA transcription[4,5], to dysregulated protein network dynamics[6-12]. While each of these may contribute individually, it is likely that a combination of these aspects is what ultimately drives disease and adds to the complexity of understanding a given disease phenotype. This 'genotype-to-phenotype' relation is one of the fundamental aspects network biologists are aiming to resolve, as, being the cellular effectors, protein signaling plays a critical role in defining the link between genotype and phenotype[7,9,12,13].

Metastasis, the spread and colonization of the primary tumor to other, often vital organs, is responsible for 90% of all cancer-related patient deaths, and thus represents the hallmark with most therapeutic potential[14]. While the development of metastases is a multistep process, requiring the acquisition of several malignant phenotypic traits[15], it is important to consider the origin of these. Extensive research has shown genomic instability to lay at the foundation of this[16-18]. However, what remains unclear is how the introduction of genomic variation leads to altered protein signaling dynamics, ultimately giving rise to the phenotypes required for metastatic progression.

Here, we have undertaken a genome-scale investigation and comparison of metastatic versus non-metastatic colorectal cancer cells, in an attempt to pinpoint specific proteins and kinases which may be fundamental to a metastatic phenotype (Figure 1). By taking an unbiased approach, both at the genome and proteome level, and deploying a multi-platform analysis including a kinome-wide RNA interference screen, we managed to identify many known and several previously unknown proteins which seem to drive the proliferation of metastatic cells specifically. To accomplish this, we have developed a computational framework for the global integration of these different datasets, and demonstrate the predictive power of our approach by validating the highlighted targets both *in vitro* and *in vivo*. Furthermore, we assess the clinical relevance of our cell-line based model by also globally assessing the genomic and proteomic landscapes of metastatic and non-metastatic patient tumors, and are investigating what percentage of metastatic colorectal cancer patients are likely to benefit from the inhibitors we tested in this study.

Results:

NGS analysis reveals different mutational landscapes

In order to characterize the impact of mutations which may play a role in the metastatic phenotype, we conducted full exome sequencing on the two cell lines using the Illumina HiSeq next-generation sequencing (NGS) platform, to determine which sequent variants are present in their genomes. In total, we identified 7,170 non-synonymous mutations, spanning 4,465 proteins in the SW480 cells, and 7,680 non-synonymous mutations covering 4,704 proteins in the SW620 cells (see Figure 2A for details). While there exists great overlap in the proteins which are affected by mutations (see Figure 2B), there is still a significant number of proteins which are uniquely altered in each cell line (~12% in SW620, ~8% in SW480), and these cell lines should consequently not be considered isogenic, despite originating from the same patient. While it is possible that some of these mutations have been acquired after establishment of the cell lines, it is likely that a significant subset of them were part of the developmental process during tumorigenesis in the patient. These data support the commonly accepted notion that cancer is a developmental disease, with the acquisition of novel mutations over time contributing to tumor malignancy[19,20]

By comparing the genomic landscapes of the two cell lines, we were able to assess the likelihood of mutations occurring. Due to the current lack of being able to predict and quantify the potential functional impact of a given mutation, an alternative method of classifying which mutations are likely to play a role in the metastatic phenotype was needed. We thus investigated whether there is a correlation between the number of mutations a certain protein has at a particular time-point (i.e. in the SW480 cells), and the likelihood of additional mutations occurring on the same protein (i.e. in the SW620 cells). For this analysis, the genomic landscape of SW480 provides the relevant background to define the expected mutational state of a protein in SW620. Thus the information content of an NGS measurement in SW620 is based on the conditional probability distribution, $P(x_{NGS,620}|x_{NGS,480})$. We construct this distribution based on two empirical observations. First, the data displays an approximate linear relation between the average number of unique mutations observed in SW620 and the number of mutations in SW480. Secondly, the distribution for the reduced quantity is well-characterized by a Weibull distribution, the shape of which is defined by only one free parameter, ν. Remarkably, as depicted in Figure 2C, the ensemble of mutational data can be summarized using only two different values of this shape parameter. In all cases, where one or more mutations have been observed in SW480, the probability of another mutation appearing in the same protein are described by $\nu \approx 0.56$. However, if no mutations have been observed in SW480, the distribution for a mutation occurring on those proteins in SW620 follows a significantly different Weibull function ($\nu_0 \approx 0.49$). These observations allow us to parametrize the required probability distribution to quantify the information content in observing a specific mutational state of a protein in SW620, and thereby use this measure as a proxy for quantifying potential involvement in metastatic progression. More details are listed in the supplementary text.

Extensive research has demonstrated kinases to be involved in many metastasis-promoting cell behaviors such as cell proliferation, migration, survival and invasion[21-29], and thereby form attractive therapeutic targets[30-33]. In fact, it has recently been shown that kinases are the most frequently mutated proteins in tumors, underlining the potential therapeutic implications they represent[34,35]. Given the diversification that tumor cells undergo throughout metastatic progression, and the ubiquitous involvement of kinase signaling in cellular signal processing[36,37], it is likely that kinase activities will be altered during this process as well. Therefore, obtaining a global overview of which kinases and other proteins are dysregulated during tumorigenesis is of great importance. From our sequencing experiments, we found that 89 and 100 kinases harbored mutations in SW480 and SW620 respectively, supporting the notion that additional kinases are affected by mutations during the process of tumorigenesis (Supplementary Table 1). However, the direct impact of these kinase-specific mutations on the disease phenotype remains to be established.

Mass Spectrometry highlights key altered proteins and phosphorylation sites in metastatic cells

In an attempt to determine the protein dynamics which may be fundamental to a metastatic phenotype, we characterized the global proteome and phospho-proteome of a matched pair of cell lines, SW480 (non-metastatic) and SW620 (metastatic), using Mass Spectrometry (MS). These cell lines originate from the same patient, where the SW480 cell line was derived from the primary adenocarcinoma in the colon, whereas the SW620 cell line was derived from a lymph-node metastasis when the cancer recurred one year later and had metastasized[38]. Importantly, when injected into the spleen of mice, the SW620 cells cause metastatic growths to form in the liver, whereas the SW480 cells do not, demonstrating the metastatic potential of the SW620 cells. The challenge is to elucidate what biochemical networks drive these different phenotypes, and additionally, considering the extensively diverse role of phosphorylation events in cellular signal processing, we also opted to look at the global phosphorylation network structure[8,9,27,36]. To more accurately mimic the in vivo micro-environment these cells would normally be exposed to, we opted to grow them on a soft layer of collagen gel instead of plastic tissue culture dishes. We hypothesize that this exposes more in vivo-relevant signaling networks as the cells are free to migrate in multiple dimensions and are receiving more natural environmental stimuli. By using SILAC labeling[39], we were able to mix the cell lysates early in the sample preparation workflow, thereby minimizing the experimental variation. Furthermore, this approach allowed us to conduct a direct comparison of protein and phosphorylation levels between the two cell lines, allowing the signaling networks to be analyzed quantitatively. To facilitate the comprehensiveness of the analysis, we used SCX fractionation, which, by reducing the sample complexity, enables greater depths of the (phospho-)proteome to be measured. Moreover, we complemented the TiO_2 enrichment, which gives rise to large numbers of phosphorylated Serine and Threonine residues[40], with phospho-Tyrosine specific enrichment using the pTyr-1000 antibody. In total, we identified 28,260 phosphorylation sites (see Figure 3C for details), of which we were able to quantify 14,699 between the two cell lines respectively. We also identified 9,070 proteins, of which 5,683 were quantifiable. Additionally, we identified 271 kinases, enabling us to use protein data on more than half the kinome in our subsequent analyses. To ensure the quality of the data, we performed the experiment both with biological and technical repeats. Biological repeats were obtained by separately labeling and growing four sets of the cell lines, whereas the technical repeats were obtained by separately performing the SCX fractionation, phospho-enrichment and sample analysis on the Mass Spectrometer. As shown in Figure 2A, the correlation between two technical repeats was very high, underlining the accurate reproducibility of the analysis. In order to determine which proteins and phosphorylation sites were significantly regulated, we analyzed the distribution of the differences between the repeats for each observed peptide. As portrayed in Figure 2B, this distribution is very narrow, allowing us to set a relatively low threshold for determining significance. In fact, by selecting the 1% and 99% quantiles, we were able to use any ratios above ±1.43, while maintaining a 1% false positive rate. In other words, the quantitative data used for the final modeling was confidently determined to be measured above the technical variation which exists in any experiment.

A current limitation in many MS based studies is the use of a common reference sequence database, which resembles the genome (and inherently, the proteome) of an individual, rather than the exact genome of the sample which is being analyzed. As we have conducted exome sequencing on the cell lines, we were able to include the genome-specific mutations in our search database, and utilize this knowledge to assess the dynamics of these mutations at the protein level where they are able to exert an effect. While further investigation is currently on-going into the exact regulation of specific mutations, we have quantitative data on 740 proteins harboring at least one non-synonymous mutation, allowing us to compare their regulation between the metastatic and non-metastatic cells.

Kinome-wide RNAi screen pinpoints different kinases to be fundamental to cell survival

In an attempt to more directly assess the role of certain kinases in maintaining the metastatic phenotype, we investigated the effect of systematically knocking down each member of the human kinome (for an exact list, see Supplementary Table 2), and subsequently measuring alterations in proliferation. Despite the genome-scale nature of the NGS and MS studies, we opted to focus on the kinome in the functional screens due to their high clinical relevance. Currently, there are about 150 small molecule kinase inhibitors under clinical investigation, and kinases are the focus of approximately 30% of all pharmaceutical research and development activities[41], underlining their therapeutic potential.

As can be seen in Figure 3D, the effect of a given kinase knockdown generally resulted in a different response in the cell lines, suggesting that the two cell lines depend on a different set of kinases for their proliferation. In this plot, the nuclei counts are normalized to 100, where 100 represents the alteration of proliferation in response to a negative siRNA (with no expected effect). Thus, any value below 100 represents decreased cellular proliferation, whereas any value above 100 represents increased proliferation. For example, the genes highlighted in blue cause a significant reduction in cell number in the SW620 cells specifically, while not having much effect in SW480; likewise, the genes highlighted in red cause a significant reduction in cell number in the SW480 cell specifically, posing potentially specific therapeutic targets in these cell lines respectively. In contrast, the genes highlighted in purple cause a significant reduction in cell number in both cell lines, and would form candidates for non-specific targets in these cell lines. Interestingly, it appears that there are far fewer genes for which the knockdown increases proliferation in SW480 compared to SW620, but an explanation for this phenomenon remains elusive. Additionally, as portrayed in Figure 3E, it appears that the SW620 cells are more resistant to phenotypic changes, as represented by the shift of the RNAi effect distribution towards a nuclei count of 100 in comparison to SW480. These results are in line with the expectations originating from the widely accepted notion that metastatic cells have obtained much greater robustness required for e.g. surviving while in circulation[14,20]. Nevertheless, these results demonstrate that specifically targeting metastatic and non-metastatic cells is a possibility, which we aim to exploit both in vitro and in vivo. To facilitate the selection of metastasis-specific targets, we deployed a novel RNAi scoring scheme, where we calculate the ratios of a given kinase knockdown based on the normalized nuclei count in SW480 and SW620. For each gene, three unique siRNAs were used, and genes that had at least two active siRNAs were taken forward for the analysis. By subsequently dividing the SW620 normalized nuclei count with the SW480 one, we derive a ratio which depicts the selectivity of a particular knockdown. As shown in Figure 3F, a knockdown with a ratio below 1 represents a gene which induces apoptosis in the SW620 cells specifically, whereas ratios above 1 represent genes which would negatively affect the SW480 growth rate specifically. This ratio can subsequently be used directly to asses the therapeutic potential of the different kinases. In this table, we are highlighting the genes that seem to be most specific for SW620.

Global Integrative Model

After generating all the data, several options were explored to facilitate a global integration, allowing to depict specific proteins which are likely to underlie the metastatic phenotype of SW620. As shown in Figure 4A, despite a significant overlap, the genes covered by each experimental approach are largely complementary. In order to avoid considerable biases by requiring a gene to have been observed in all three datasets, and rather, utilize the strength of the individual approaches, it was opted to integrate them based on information theoretical measures. The method is defined in depth in the supplementary material, and a conceptual overview is illustrated in Figure 4B. In principle, the approach relies on where a given datapoint falls within its respective distribution, and how much information is therefore contained in the measurement. We term this information 'energy'. We include and integrate the following measures: the protein ratio and phosphorylation ratio from the phospho-proteomics experiments, the nuclei ratio from the RNAi screen and the genomic information from the sequencing experiments. To account for missing values, we utilize the expected information as defined by the distributions of the individual datasets, and include an 'entropy' term, which allows setting the contributed information content of a particular dataset to 0 in case of a protein not having been observed. The terms originating from the individual datasets can subsequently be summed to represent the total 'energy' of a specific protein, which allows for a ranked hit list to be established (Figure 4C).

In vitro validation

After compiling our final target list, we set out to confirm whether the modulation of these targets using available small molecule inhibitors would have the expected effect (Figure 4D). To this end, we first conducted dose-response experiments, to determine appropriate dosing concentrations and drug effectiveness. As shown in Figure 5A, we ranged the dose from 100uM to 0.3nM, and determined each drug's IC50 value. What is evident however, is that 1) not every drug was able to induce a significant reduction in cell number, and 2) not every drug had a physiologically relevant IC50 value. Only in the case of Dasatanib, Dovitinib, Foretinib, HG-9-91-01, Motesanib, MRT199665, Sunitinib, TAE684 and XL184 were we able to successfully establish EC50 values. Additionally, the difference in response between SW480 and SW620 was not always significant, indicating that not all drugs were able to specifically target the metastatic cells. Nevertheless, these experiments highlighted Sunitinib as a potential metastasis-specific drug, as a maximum response in SW620 was achieved at a dose of 10uM, at which point the effect in SW480 had not yet been saturated. In fact, even at a dose of 100uM, we still did not achieve a full response in SW480, hence suggesting a lack of full activity of this compound in this cell line. SW620 on the other hand, had an EC50 value of 3.6uM, making it a highly relevant drug for follow-up validation. Additionally, TAE684 also displayed a significant response, and was selected for follow-up experiments, as was Dasatanib, Foretinib, Motesanib, HG-9-91-01, MRT199665 and PP2. This last inhibitor was selected despite not showing a reasonable response, as it is the most specific inhibitor available for Fyn kinase, one of the top hits from our modeling approach.

Once it was established which inhibitors were able to induce a response when applied in isolation, we attempted to determine whether we could improve the effectiveness of Sunitinib by combining it with other inhibitors. As shown in Figure 5B, we not only tested inhibitors in combinations, but also attempted to conduct combination treatment in a time-staggered fashion, as inspired by recent work by Lee and colleagues[28]. We depict the effect of both the single combination and the time-staggered combination treatment (pre-treatment for 24hrs, additional treatment for another 24hrs), but these inhibitor combinations did not seem to benefit from such a strategy. Nevertheless, the combination treatments did appear to have a beneficial effect, with increased apoptosis being observed in all the combinations except PP2 and HG-9-91-01.

Patient sample analysis

In an attempt to assess the potential clinical relevance of our cell line-based findings, we undertook a global study of the genomic and (phospho-)proteomic profiles in a panel of patient samples. Samples originating from the four different Dukes' stages (I - IV) were obtained, representing the tumorigenesis from non-metastatic to fully metastatic disease. The proteomics experiments were conducted quantitatively, with the SW480 cells being labeled with a medium label, the SW620 cells with a heavy label and the patient samples naturally representing the light label. In this manner we were able to directly quantify and compare the protein and phosphorylation levels to the two cell lines, and assess how representative they are when compared to a general patient population. As listed in Figure 6A, we managed to identify several thousands of non-synonymous mutations, proteins and phosphorylation sites across these samples, thereby representing a rich source of data for trying to decipher the genomic and proteomic players which define the disease phenotype in these patients. While extensive analysis is currently still on-going, it becomes evident from Figure 6B that patient heterogeneity is a severe obstacle in this analysis. Despite a large amount of proteins having been observed and quantified in all the samples (1634), it appears that they are not sufficient for classifying which stage a given tumor belongs to, as evidenced by a lack of clustering patterns. Similarly for the phosphorylation sites which were quantified across all 12 samples, there appears to be a lack of clear clustering patterns, thereby suggesting that the phenotype establishment in these patients is driven by a more complex interplay of these proteins, phosphorylation sites and mutations in their respective signaling network states. An in-depth analysis into this aspect is currently on-going, which will likely reveal which critical network attractors define the disease phenotype.

Mouse data

After the initial in vitro screens, we aimed to validate the relevance of the successful inhibitors in an in vivo setting. Both single and combination treatments were tested in a subcutaneous tumor implantation model. As shown in Figure 6C, from the single treatments, Sunitinib and Foretinib were able to reduce tumor growth significantly (Foretinib: p= 0.0065**, Sunitinib: p= 0.0006***) compared to vehicle treatment, whereas Dasatanib, Motesanib and TAE684 were unable to do so (Dasatanib: p= 0.4009, Motesanib: p= 0.2196, TAE684: p= 0.7705). P values were calculated from a linear regression analysis of the difference between treatment and vehicle. In the combination treatments (Figure 6C), all combinations demonstrated a significant reduction (Sunitinib + Dasatanib: p= 0.0098**, Sunitinib + Motesanib: p= 0.0014**, Sunitinib + Foretinib: p= 0.0034**, Sunitinib + TAE 684: p= 0.0009***), but none displayed greater effectiveness than Sunitinib alone (detailed in Figure 6D). This is in line with the effect of the single treatments, but it remains to be seen if there are beneficial effects in the longer term as experiments are still on-going. Data for the SW480 cells is also being generated, but currently unavailable as the subcutaneous tumors take much longer to establish. Based on the currently available data, it is planned to take Sunitinib and Foretinib forward into an intrasplenic metastasis assay, both as single and combination treatments, in order to appropriately assess their effectiveness in preventing metastatic tumor formation.

Discussion:

In this study, we have demonstrated the power of combining several technological platforms to assess different biological aspects, in an attempt to shed light on the roles and complex interplay of the genomic and proteomic levels of signal processing by the cell in ultimately deciding a phenotypic response. Our approach enables one to globally study the protein and phosphorylation dynamics, and to pursue the discovery of how genomic alterations may affect these. Additionally, by using genome-specific search databases originating from the NGS experiments, we are able to monitor how mutations at the genome level are propagated to the proteome level and interrogate their dynamic modulation. This allows one to start deciphering how these mutations may be utilized by the cell for determining its phenotype. Despite the relatively low number of different mutations between the two cell lines, a high number of proteins and phosphorylation sites were detected with altered expression

levels. This seems to highlight the role of the proteome dynamics in establishing a particular phenotype. Furthermore, the value not only lies in the overlap of the approaches, but also in their complementarity, as a potential therapeutic candidate may not have been observed in all three of the datasets. For example, a gene harboring a specific mutation may play a critical role in the disease development, but may simply not have been observed using MS due to its inherent dynamic range limitation. MYO3A, MYLK3, WNK3 and PRKG2 (see Figure 4C) are good examples of these, as we only had genomic and phenotypic RNAi data on these kinases. Therefore, if MS had been deployed in isolation, they would not have been detected. Similarly, for e.g. Fyn, DCLK1 and TNIK, while we did not observe any mutations, we did observe a significant increase in protein expression and a significant decrease in cell number in the SW620 cells upon knockdown, thereby strongly suggesting their implication in maintaining SW620 cell viability. This would not have been detected through the use of NGS alone. Thus, an additional strength of this integrative approach is clearly signified by the use of an RNAi screen. By directly assessing the role of a particular gene in a phenotype of interest (in this case, the rate of proliferation of metastatic versus non-metastatic cells), we were able to extend greater relevance to a given MS or NGS observation. It is likely that, had we deployed a genome-wide RNAi screen, even more targets can be found, but we opted against this due to the general lack of well-characterized small molecule inhibitors to non-kinase protein targets. Additionally, as exemplified by the in vitro screen results, the inherent lack of specificity of small molecule inhibitors seems to hamper the direct translation of an RNAi-based observation to a clinical inhibitor. While some genes showed great promise in the RNAi screen, some of the inhibitors covering these kinases in their target-spectra failed to reproduce those results. Hence why it was preferable to use well-studied inhibitors with known kinase-specific inhibition profiles in this study[42].

Based on both the in vitro and in vivo validation studies that were undertaken, Sunitinib seems to be a promising drug for treatment of metastatic colorectal cancer. It has been shown to have a broad target spectrum, to which some if its success may be attributed[42]. From the targets highlighted by our analysis, Sunitinib displayed a sub-micromolar affinity to 22 of them. Interestingly, Pfizer has also investigated its clinical potential in several clinical trials[43,44]. A Phase 3 clinical trial was halted prematurely, mainly due to toxicity effects, and we argue this was likely due to both inefficient patient selection and non-optimal treatment combination with current standard-of-care. For the former, it would be critical to assess the protein network signatures of the tumors to assess the likelihood of response to Sunitinib; for the latter, the combination of a broad spectrum kinase inhibitor with fluorouracil, leucovorin, and irinotecan seems to have given rise to additional toxicity-related adverse events and lack of concomitant response improvement [43]. In the cell lines, we detected 76 Sunitinib protein targets, of which 48 were up-regulated in SW620 and 28 were up-regulated in SW480, providing a potential explanation for the positive response in SW620. In the patient samples analyzed in this study, on average, 34 target proteins of Sunitinib were detected, of which 15 displayed increased levels of expression compared to SW480. There was no clear visible trend with regard to more of these proteins to be up-regulated in later stages of tumor development however, again underlining the importance of assessing the dynamic proteomes of tumors before prescribing treatment. When comparing the cell lines proteomic profiles with the patient samples, a potential reason for the apparent lack of overlap could be attributed due to it being the primary tumor which is analyzed. While the primary tumor originating from the later tumor stages is known to have metastasized, the actual subpopulation within the primary tumor which is metastatic may however be rendered undetectable. One reason why SW620 is a powerful metastasis model system is due to its origin from a metastatic site, and we argue that one would therefore need to look at distant metastases in the patient as well, in order to gain a better understanding of the molecular networks in these. Nevertheless, our results indicate that the disease phenotype establishment in these patients is likely driven by a more complex interplay of their proteins, phosphorylation sites and mutational alterations than our initial analyses were able to capture. To this end, integration of publicly available datasets on the genomic landscapes of colorectal tumors will likely be vital, as it greatly extends the number of observations we can include in our model. Especially bearing in mind the accumulation of mutations throughout cancer development, it will be critical to assess a larger population of tumors, as based on

the numbers alone (Figure 6A,B), there does not appear to be a clear pattern in the samples we analyzed.

In conclusion, while additional analysis and experimental work is currently still on-going, our work thus far has demonstrated the value of generating and integrating several types of datasets, in order to assess different molecular mechanisms in depth. By taking a global approach, without much *a priori* knowledge, and utilizing a novel unbiased integrative algorithm, we were able to pinpoint specific proteins which may be fundamental to metastatic survival, and validate our findings in vitro and in vivo. We demonstrate the importance of conducting in vivo validation, as not all our in vitro results were confirmed in vivo, thereby narrowing down the treatment options to potentially more clinically relevant ones. While not all the combination treatments were successful, a potential benefit may lie in being able to reduce doses for treatment, thereby reducing treatment side-effects. Additional metastasis-specific in vivo models are currently in progress, with the intra-splenic method being adopted to assess the efficacy of the described inhibitors in preventing metastatic growths. Overall, despite the seemingly infinite complexity of cancer, by focusing on one particular aspect and analyzing it in-depth, potential clinical benefit may be within reach.

Acknowledgements:
The authors would like to thank members of the Linding and Erler labs for useful input to the manuscript. Additionally, we would like to thank Chiara Francavilla and Jesper V. Olsen (CPR, Denmark) for technical assistance with the phospho-proteomics experiments, and to Agata Wesolowska-Andersen for help with processing the cell line NGS data. We would also like to thank Antonio Palmeri for bioinformatics assistance with domain mapping.

Figure legends:

FIG.1 Conceptual workflow of the study

Outline of the analysis pipeline workflow, highlighting the individual analyses to be done (exome sequencing and (phospho-)proteomics on both the cell lines and patient samples, and a kinome-wide RNAi screen measuring cell proliferation. This data is subsequently computationally integrated to result in a ranked hit list which are fundamental to metastatic cell survival. These hits are finally functionally validated in vitro and in vivo.

FIG.2 Next-Generation Sequencing Data

A) Table of number of mutations, mutated proteins and mutated kinases in the cell lines

B) Venn diagram portraying the overlap of the mutated proteins between the two cell lines. Despite originating from the same patient, these are clearly not isogenic due to 12% and 8% unique proteins harboring a mutation in SW620 and SW480 respectively.

C) Weibull function plot of mutational landscape in SW620. In all cases, where one or more mutations have been previously observed in SW480, the probability of another mutation appearing in the same protein are described by $\nu \approx 0.56$ (solid line). However, if no mutations have been observed in SW480, the distribution for a mutation occurring on those proteins in SW620 follows a significantly different Weibull function ($\nu_0 \approx 0.49$, dashed line). This measure is subsequently used as a proxy for quantifying potential involvement in metastatic progression.

FIG.3

A) Scatter plot showing the consistency of measurements over the two technical replicates

B) Density plot highlighting the distribution of errors between two technical replicates. The quantiles are listed to show the ratio cutoffs for specific false positive rates.

C) Statistics from the mass spectrometry screen that was done on the cell lines.

D) Scatter plot showing the normalized nuclei count in SW480 and SW620 for a specific kinase knockdown over the complete kinome screen. Genes specifically affecting SW620 and SW480 are highlighted in blue and red respectively, whereas genes affecting both the cell lines are highlighted in purple. The markers at 100 indicates no effect.

E) Density plot for visualizing the kinome-wide RNAi effects in both the cell lines. It is evident that SW480 is more susceptible to apoptosis induction, whereas SW620 is more resilient. The marker at 100 indicates no effect.

F) Table highlighting the top genes which significantly affect cell number specifically in SW620. The ratio, derived from dividing the normalized nuclei count in SW480 by SW620 is a measure of SW620-specific therapeutic potential.

FIG.4

A) Overlap of genes covered by the respective experimental approaches.

B) Conceptual overview of global data integration model, where the total information content of each measurement is integrated to determine the global 'energy' of a given protein.

C) The top hit list after the global data integration. Rows highlighted in green are covered by the inhibitors we selected for functional validation, yellow rows are proteins for which there were no available inhibitors and rows presented in red were tested unsuccessfully.

D) Overview of kD / IC-50 values of the compounds for the top targets highlighted by the energy model. For inhibitors marked with *, values represent the remaining kinase activity at 1μM.

FIG.5

A) 12-point dose response curves for the selected inhibitors in both SW480 and SW620.

B) Combination and time-staggered treatment results of successful inhibitors, with results being highlighted for both cell lines.

FIG.6

A) Table presenting details on the mutation and proteomics data generated from the patient samples. Patients A through D represent Dukes' Stage I-IV respectively, and we analyzed 3 samples per stage.

B) Heatmaps highlighting lack of clustering between tumor stages, thereby suggesting that the tumor heterogeneity cannot be captured by the proteins and phosphorylation sites detected in a simple manner. Current analysis into this problem is still on-going.

C) In vivo growth curves for single and combination treatments in mice implanted subcutaneously with SW620, highlighting Sunitinib and Foretinib as potent metastasis treatments. SW480 data is currently being generated.

D) Detailed in vivo growth curves for the combination treatments, highlighting a lack of significant improvement over single Sunitinib treatment.

Methods & Materials:

Sample preparation for Cell-line (phospho-)proteomics. SW480 and SW620 cells (regularly checked for mycoplasma contamination) were SILAC labeled over six passages, after which label incorporation was determined to be > 95%. Labeling was done as follows: two initial vials of SW480 and SW620 were labeled heavy, and two initial vials of SW480 and SW620 were labeled light, in order to allow for a label-swap experiment and resulting in 4 biological replicates. The serum-starved cells were subsequently seeded at ~80% confluency on 1.25mg/ml collagen gel coated 15cm dishes, and left to settle for 24hrs. Cells were lysed with ice-cold modified RIPA buffer supplemented with 4M Urea, Roche complete protease inhibitor cocktail tablets and ß-glycerophosphate (5mM), NaF (5mM), Na-orthovanadate (1mM, activated). Lysates were sonicated on ice and spun down at 4,400xg for 20mins at 4°C. Proteins were precipitated over-night in ice cold Acetone at -20°C, and dissolved in 6M Urea, 2M Thiourea, 10mM HEPES pH 8.0 at room temperature (RT). Protein concentrations were determined using Bradford, and heavy and light samples were mixed 1:1 (12mg each). Subsequently, the samples were reduced with 1mM DTT for 1hr, and alkylated with 5.5mM Chloroacetamide for 1hr, after which they were pre-digested with Lysyl Endopeptidase (Wako) at a 1:200 enzyme-to-protein ratio for 4hrs at RT. Lysates were diluted 1:4 with 50mM Ammonium Bicarbonate, after which Trypsin (MS grade, Sigma) was added at a 1:200 enzyme-to-protein ratio and left rotating over-night at RT. Enzymatic activity was quenched by adding TFA to a final concentration of 2%, after which the samples were clarified by spinning down at 2,000xg for 5 minutes and desalted using 360mg SepPak columns (Waters WAT020515). Peptides were eluted using 2x 2mL of 40% AcN, 0.1% TFA, and 1x 2ml of 60% Acetonitrile, 0.1% TFA.

For the global, Titanium Dioxide (TiO2) based phospho peptide enrichment, the eluent was directly subjected to SCX fractionation, where peptides were separated over a 0-30% Buffer B gradient in 60 minutes at a 1ml/min flowrate (Buffer A: 5mM potassium dihydrogen phosphate, 30% Acetonitrile, pH2.7; Buffer B: 5mM potassium dihydrogen phosphate, 30% Acetonitrile, 350mM potassium chloride, pH2.7). The resulting fractions were pooled according to their chromatography into 11 final samples, which were enriched seperately for phosphorylated peptides. Six aliquots were taken at this point for the global proteome analysis. The TiO2 enrichment was conducted similarly to [45], with several adjustments. For the TiO2 loading solution, 0.02g/ml dihydrobenzoic acid was dissolved in 30% Acetonitrile and 4% TFA, and the TiO2 beads were incubated in this solution for 15 minutes prior to peptide enrichment. Each pooled SCX fraction was enriched with 1.7mg of TiO2 beads suspended in 6ul of TiO2 loading solution, and left to rotate end-over-end for 30 minutes at RT. The flow-through (early eluting fractions) was enriched three times consecutively, whereas the single SCX chromatography peak peptide samples were enriched twice. Samples were spun at 2000xg for 5 minutes (RT), and pelleted beads were washed with 100ul SCX Buffer B. Subsequently, beads were pelleted again (2000xg, 5minutes, RT) and washed with 100ul 40% Acetonitrile, 0.25% acetic acid, 0.5% TFA. Finally, pelleted beads were re-suspended in 50ul 80% Acetonitrile, 0.5% acetic acid, and transferred to separate in-house packed C8 StageTips[46]. Liquid was spun through at 3000 rpm for 1 minute, after which the phosphorylated peptides were eluted with 1x 20ul 5% Ammonia and 1x 20ul 10% Ammonia, 25% Acetonitrile into a 96-well PCR plate, containing 20ul of 1% TFA, 5% Acetonitrile solution. Peptides were concentrated to a total volume of 10ul in an Eppendorf Speedvac, and acidified with 40ul of 1% TFA, 5% Acetonitrile, after which they were desalted on in-house packed C18 StageTips prior to LC-MS analysis.

For the pTyr specific phospho peptide enrichment, the SepPak eluent (equating to 24mg of peptides) was concentrated in an Eppendorf Speedvac and stored at -80°C. pTyr specific enrichment was conducted with the pTyr-1000 antibody from CST, using the protocols provided by the manufacturer, and the samples were run as technical duplicates on the LC-MS.

Sample preparation for patient sample (phospho-)proteomics.

For the patient sample analysis, SW480 cells were labeled with a medium SILAC label and the SW620 cells were labeled heavy. After determining >95% label incorporation, the cells were lysed,

ammonia precipitated and dissolved in 6M Urea, 2M Thiourea and 10mM HEPES pH 8.0 as described above. Patient samples were obtained from the Aberdeen Tissue repository, following all applicable ethical guidelines. The flash-frozen patient samples underwent denaturation in a Denator instrument to prevent any protein activity during processing, after which they were lysed in 6M Urea, 2M Thiourea and 10mM HEPES pH 8.0 using a Qiagen TissueRuptor. The lysates were sonicated on ice and spun down at 4,400xg for 20mins at RT. Protein concentrations of tissue and cell lysates were measured using Bradford, and mixed in a 1:1:1 ratio at 8mg each. The samples were reduced with 1mM DTT for 1hr, and alkylated with 5.5mM Chloroacetamide for 1hr, after which they were pre-digested with Lysyl Endopeptidase (Wako) at a 1:200 enzyme-to-protein ratio for 4hrs at RT. Lysates were diluted 1:4 with 50mM Ammonium Bicarbonate, after which Trypsin (MS grade, Sigma) was added at a 1:200 enzyme-to-protein ratio and left rotating over-night at RT. Enzymatic activity was quenched by adding TFA to a final concentration of 2%, after which the samples were clarified by spinning down at 2,000xg for 5 minutes and desalted using 360mg SepPak columns (Waters WAT020515). Peptides were eluted using 2x 2mL of 40% AcN, 0.1% TFA, and 1x 2ml of 60% Acetonitrile, 0.1% TFA. The eluent was concentrated in an Eppendorf Speedvac and stored at -80C. Due to the limited amount of protein available from the tissue samples, pTyr and TiO2 enrichment was conducted sequentially, where the dried-down peptides were subjected to pTyr enrichment according to manufacturers instructions with the pTyr-1000 antibody, after which the supernatant was diluted 1:10 with 0.1% TFA, made up to 2% TFA and purified on a SepPak again for subsequent TiO2 enrichment and proteome analysis as described above.

LC-MS Analysis

For the cell line LC-MS analysis, peptides were eluted from the StageTip with 2x 20ul 80% Acetonitrile, 0.1% Formic acid, and concentrated to 5ul final volume. The eluent was acidified with 1% TFA, 2% Acetonitrile and loaded onto a 50cm C18 EasySpray column (ThermoFisher, ES803), using the Thermo EasyLC 1000 uHPLC system and the column oven operating at 45°C. Peptides were eluted over a 250 minute gradient, ranging from 6-60% of 80% Acetonitrile, 0.1% Formic acid, and a Q Exactive (ThermoFisher) was run in a DDA-MS2 top10 method. Full MS spectra were collected at a resolution of 70,000, with an AGC target of 3e6 or maximum injection time of 20ms and a scan range of 300-1750 m/z. The MS2 spectra were obtained at a resolution of 17,500, with an AGC target value of 1e6 or maximum injection time of 80ms. Dynamic exclusion was set to 20s, and ions with a charge state < 2 or unknown were excluded. For the proteome samples, the settings were the same, except for a gradient time of 240mins, maximum MS2 injection time of 60ms and dynamic exclusion of 45s.

For the patient sample LC-MS analysis, peptides were eluted from the StageTip with 2x 20ul 80% Acetonitrile, 0.1% Formic acid, and concentrated to 5ul final volume. The eluent was acidified with 1% TFA, 2% Acetonitrile and loaded onto a 50cm C18 EasySpray column (ThermoFisher, ES803), using the Thermo EasyLC 1000 uHPLC system and the column oven operating at 45°C. Peptides were eluted over a 240 minute gradient, ranging from 6-60% of 80% Acetonitrile, 0.1% Formic acid, and an Orbitrap Fusion (ThermoFisher) was run in a DDA-MS2 top speed method with a 3s cycle time, and both HCD and ETD fragmentation was being deployed depending on the precursor peptide charge state and mass. This decision-tree method was based on [47]. MS spectra were acquired in the Orbitrap at 120,000 resolution with a maximum IT of 20ms or AGC target value of 4e5. MS2 spectra were acquired in the Ion trap at 30,000 resolution with a maximum IT of either 80/120, 100/100 or 200/200 ms (HCD/ETD) or AGC target value of 1e4. Precursor isolation window was set to 1.6Da, collision energy at 35%, and dynamic exclusion at 60s.

Computational analysis of MS data.

The resulting raw files were searched in MaxQuant Version 1.2.7.4 for the cell-line experiments, and MaxQuant 1.4 for the patient samples (due to lack of Orbitrap Fusion support in older version). All searches were conducted on SW480/SW620 specific database, where all protein sequence variants were included in addition to the wild-type Ensembl v.68 human FASTA sequences. The raw data have

been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[48] with the dataset identifier XXX. Variable modifications were set as Methionine oxidation, Protein N-term acetylation and Serine/Threonine/Tyrosine phosphorylation, and Cysteine carbamidomethylation was set as a fixed modification. FDR rates were set to 1%, and the 'match between runs' functionality was activated.
Results from the searches were stored in a MySQL database, and all further analysis was done using scripts written in-house on our "CoreFlow" platform, based on the R statistical package, MySQL and Python. All code and data will be released to the public upon request. Phospho-peptide search results filtering was based on phosphorylation localization probability >= 0.75, minimum MaxQuant peptide ID score of 50 and a minimum number of unique MS observations of 3, in order to only use high confidence identifications. For determining quantitative protein ratios, for each protein we used the mean ratio of 3 unique peptides (with MaxQuant peptide ID score >=50) without any modifications.

Sample preparation for sequencing and data analysis. SW480 and SW620 cells were grown to 80% confluency in a T-75 flask, and DNA extraction was performed using reagents and instructions provided with the Qiagen QIAamp DNA Mini kit. 5 ug of purified DNA were sent to Roche Nimblegen for full exome sequencing using the SeqCap EZ Human Exome Library v3.0 capture kit. High-quality reads, with > 80x mean coverage and > 95% of exome bases at 10x coverage, were obtained from sequencing and aligned to the NCBI37 reference human genome (version GRCh37) using the Burrows–Wheeler Alignment Tool. The alignment was refined by means of quality score recalibration and around indel realignment using Genome Analysis ToolKit package. SNP calling was performed with SAMtools package using default settings. Next, results were further filtered with VCFtools using standard default settings as well as a minimum 10x sequencing depth threshold set for SNP calling. The data was further analyzed with the help of SAMtools and BEDtools packages and custom-written Perl and Python scripts. Finally, fasta files for both wild-type and mutant protein sequences were generated using the Variant Effector Predictor (VEP) package from Ensembl.

For the patient samples, 25mg of tumor tissue was subjected to DNA extraction using reagents and instructions provided with the Qiagen QIAamp DNA Mini kit. 5 ug of purified DNA were sent to BeckmanCoulter Genomics for full exome sequencing using the SeqCap EZ Human Exome Library v3.0 capture kit and the Illumina HiSeq platform. The paired-end reads from Illumina HiSeq were mapped to the HG19 reference genome using BWA. Then Picard-tools was used to sort the output BAM file and mark duplicates. Local indel realignment and base quality score recalibration was done with GATK. The aligned reads were filtered based on mapping quality score using Samtools (MAPQ >= 30). GATK UnifiedGenotyper was used to detect variants. The final variant filtering was done using the VCF-annotate tool and a Python script developed in-house (read depth >= 10, root mean square mapping quality score >= 55).

Kinome-wide RNAi screen:
Cells were transfected with Silencer siRNAs (Life Technologies) using a 'one-step' method; siRNAs were diluted to 500nM in OptiMEM (Life Technologies) and mixed 1:1 with Lipofectamine RNAiMAX, also diluted in OptiMEM, such that each siRNA was mixed with 0.06µl of reagent for SW480 transfection and 0.08µl of reagent for SW620 transfection. The siRNA/transfection reagent mix was then incubated at room temperature for 15 minutes prior to being dispensed into collagen-coated CellCarrier (PerkinElmer) plates. SW480 and SW620 cells were plated directly into the siRNA containing wells at a density of 4000 cells per well. Cells were then incubated with the siRNAs for 72 hours at 37°C, 5% CO2, 95% humidity before being fixed, stained and read on the Opera High Content Imaging reader (PerkinElmer). siRNAs were diluted 1 in 10 by the addition of cell culture medium giving a final, 'in-assay' concentration of 50nM.

Cells were fixed by the addition of 4% paraformaldehyde (Sigma), incubated at room temperature for 15 minutes. Parafromaldehyde was then removed and cells were stained with Hoechst 34580 (Life Technologies) diluted to 2µg/ml in PBS. Cells were incubated for 1 hour, at room temperature, in the

dark before being washed and imaged. Cells were imaged on the Opera using a x20 water objective, 405nm laser excitation and 450/50 band pass emission filter. Nuclei were detected using the Acapella image analysis software (PerkinElmer).

In vitro target validation screen:
Cells were plated at a density of 4000 cells per well and incubated overnight at 37°C, 5% CO2, 95% humidity. Compounds were diluted in PBS and then added to cells at a 1 in 5 dilution. Cells were incubated with compounds for 48 hours (at 37°C, 5% CO2, 95% humidity) before being fixed with 4% paraformaldehyde, stained with Hoechst 34580 and imaged, as described previously.

Primary Subcutaneous Tumor Model.
Adult female immunodeficient CD1 nude mice (Charles River/SCANBUR, Denmark), 8 weeks old and weighing 22–27 g, were injected subcutaneously into the flank with either luciferase-expressing SW480
(4×10^6 cells) or SW620 (2×10^6 cells) resuspended in 100 µL Hanks Balanced Salt Solution (HBSS), using a 1-mL syringe and 30-gauge needle (n=4 tumours per treatment group, n=8 tumours vehicle). Mice
were then randomised into ten treatment groups. Single treatment groups included; Vehicle (2.5% DMSO, 80mM Sodium Citrate), Dasatanib (15 mg/kg), TAE684 (10mg/kg), Foretinib (100mg/kg), Motesanib
(100mg/kg) and Sunitinib (60mg/kg). Combination treatments included Sunitinib + Dasatanib, Sunitinib + TAE684, Sunitinib + Foretinib and Sunitinib + Motesanib at dosing concentrations stated above. Treatment involved daily oral gavage at stated doses (in 200uL volume) once palpable tumour size had reached $4mm^3$. Inhibitor stocks were solubilised in DMSO, aliquoted and stored at -80°C. Stock solutions
were diluted to working concentration fresh daily. Treatments continued for 4 weeks or until tumours reached a maximum allowed volume (0.90 cm^3). Tumour volume and body weight of all mice were
measured twice weekly using callipers and scales respectively. All in vivo experiments were under authorization and guidance from the Danish Inspectorate for Animal Experimentation according to guidelines for the welfare and use of animals in cancer research.

Intrasplenic implantation Metastatic Model.
Adult female immunodeficient CD1 nude mice (Charles River/SCANBUR, Denmark), 8 weeks old and weighing 22–27 g, were anaesthetized (1:1:3 Hypnorm:Hypnovel:water); dose, 10 mL/kg). A small incision was made on the left side of the abdomen, and the spleen was exposed. Mice were then injected into the spleen with either luciferase-expressing SW480 or SW620 cells (2×10^6 cells per mouse per cell line; n = 8 mice per cell line), resuspended in 50 µL Hanks Balanced Salt Solution (HBSS), using a 1 mL insulin syringe, and 30-gauge needle. Mice were then randomly divided into treatment groups. Oral gavage of these mice with began on day 7 and was continued for 4-5 weeks. Once weekly, mice were
injected with 120 mg/kg luciferin and metastatic dissemination of the cells was monitored using IVIS Lumina II (Caliper Lifesciences, Runcorn, UK). Metastasis was monitored weekly and quantified by
measuring luminescent signal from each organ at the experimental endpoint. All in vivo experiments were under authorization and guidance from the Danish Inspectorate for Animal Experimentation according to guidelines for the welfare and use of animals in cancer research.

Global data integration:

The data has been integrated based on information theoretical measures. Let x_d be the measurement of a given protein with respect to data of type d, representing either the protein ratio ($d = \text{mass}$) or phosphorylation ratio ($d = \text{ph}$) from the phospho-proteomics experiment, the nuclei ratio from the RNAi screen ($d = \text{RNAi}$) or the sequential information ($d = \text{ngs}$) from the sequencing. The information from the phospho-proteomics ($d =$mass,$d =$ph) is collectively refered to as MS in figure 4. Specifically, x_{mass} and x_{ph} represent the log-ratio of the peak-intensities between SW620 and SW480 for the protein concentration or phosphorylation concentration respectively. The overall phosphorylation state of a given protein is summarized in terms of the minimum, median and maximum ratios observed for all the different phosphorylated sites, so $x_{\text{ph}} = (x_{\text{ph}-\text{min}}, x_{\text{ph}-\text{median}}, x_{\text{ph}-\text{max}})$. x_{RNAi}represents the ratio of number of nuclei in SW620 to SW480 as identified by the imaging analysis of the RNAi screen. In all three cases, we use the median over biological repeats to define x. For the sequencing part, we summarize the information in terms of the number of mutations observed for a given protein in SW480, $x_{\text{ngs},480}$ and the number of unique mutations observed in SW620, $x_{\text{ngs},620}$.

We quantify the relevance of a protein in defining the metastatic state of the cell, by the level of 'surprise' viz. information obtained from the protein measurement in SW620 in comparison to the measurement of SW480. The information pertaining to a given observation is defined as $I(x_d) = -\log(P(x_d))$, where $P(x_d)$ is the probability of x_d. For the phospho-proteomics data and RNAi data we simply construct $P(x_d)$ from the actual set of observed values over all proteins using kernel estimation. For the mutational analysis, the sequencing of the SW480 provides the relevant background to define the expected mutational state of a protein in SW620. Thus the information content of a sequencing measurement in SW620 is based on the conditional probability distribution, $P(x_{\text{ngs},620} \,|\, x_{\text{ngs},480})$. We construct this distribution based on two empirical observations. First, the data displays an approximate linear relation between the standard deviation, σ, of $x_{\text{ngs},620}$ and the number of mutations in SW480, ie. $\sigma = \sigma(x_{\text{ngs},480}) \approx a_0 + a_1 \cdot x_{\text{ngs},480}$. Secondly, the distribution for the reduced quantity, $\tilde{x} = \frac{x_{\text{ngs},620}}{\sigma}$, is well-characterized by a *Weibull function,* $f_{WB}(\tilde{x} \,|\, \nu) = \frac{\nu}{\lambda}\left(\frac{\tilde{x}}{\lambda}\right)^{\nu-1} \exp\left(-\left(\frac{\tilde{x}}{\lambda}\right)^{\nu}\right)$, where λ represents a characteristic scale and ν represents the shape. Note that for any given choise of ν, $\lambda = \lambda(\nu)$ is fixed from the requirement that the standard deviation of $\tilde{x}$ should be unity. Remarkably, the ensemble of mutational data can be summarized using only two different values of the shape parameter. In all cases, where one or more mutations have been observed in SW480, $P(\tilde{x})$ is given by $P(\tilde{x}) \approx f_{WB}(\tilde{x} \,|\, \nu = 0.56)$, whereas the distribution for $\tilde{x}$ has a significantly different Weibull shape, $P_0(\tilde{x}) \approx f_{WB}(\tilde{x} \,|\, \nu_0 = 0.49)$ when $x_{\text{ngs},480} = 0$. These observations allow us to parametrize the required probability distribution, $P(x_{\text{ngs}}) = P(x_{\text{ngs},620} \,|\, x_{\text{ngs},480}) = f_{WB}\left(\frac{x_{\text{ngs},620}}{\sigma(x_{\text{ngs},480})}\right)$, to quantify the information content in observing a specific mutational state of a protein in SW620.

The information content pertaining to the combined observation, $x = (x_{\mathrm{mass}}, x_{\mathrm{ph}}, x_{\mathrm{RNAi}}, x_{\mathrm{ngs}})$ is now given as $I(x) = \sum_d I(x_d)$. In order to have equal weighting over the complementary types of data-information we set $I(x_{\mathrm{ph}}) = \frac{1}{3}(I(x_{\mathrm{ph-min}}) + I(x_{\mathrm{ph-median}}) + I(x_{\mathrm{ph-max}}))$. Furthermore, to account for missing data we use the expected information (viz. entropy), $\langle I(x_d)\rangle = \int I(x_d)P(x_d)dx_d$, for each x_d as baseline, so that the final information score ('energy') is defined as $E(x) = \sum_d E(x_d)$, where $E(x_d) = I(x_d) - \langle I(x_d)\rangle$. This implies that the 'energy' of a missing component of x can simply be set to zero, corresponding to the expected 'energy' over $P(x_d)$.

Bibliography

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
2. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719-724 (2009).
3. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546-1558 (2013).
4. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27 (2012).
5. Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nat Rev Cancer* **4**, 143-153 (2004).
6. Cox, J. & Mann, M. Is proteomics the new genomics? *Cell* **130**, 395-398 (2007).
7. Creixell, P., Schoof, E. M., Erler, J. T. & Linding, R. Navigating cancer network attractors for tumor-specific therapy. *Nat Biotechnol* **30**, 842-848 (2012).
8. Pawson, T. & Hunter, T. Signal transduction and growth control in normal and cancer cells. *Curr Opin Genet Dev* **4**, 1-4 (1994).
9. Pawson, T. & Linding, R. Network medicine. *FEBS Lett* **582**, 1266-1270 (2008).
10. Pawson, T. & Kofler, M. Kinome signaling through regulated protein-protein interactions in normal and cancer cells. *Curr Opin Cell Biol* **21**, 147-153 (2009).
11. Taylor, I. W. & Wrana, J. L. Protein interaction networks in medicine and disease. *Proteomics* **12**, 1706-1716 (2012).
12. Vidal, M., Cusick, M. E. & Barabasi, A. L. Interactome networks and human disease. *Cell* **144**, 986-998 (2011).
13. Gstaiger, M. & Aebersold, R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet* **10**, 617-627 (2009).
14. Gupta, G. P. & Massague, J. Cancer metastasis: building a framework. *Cell* **127**, 679-695 (2006).
15. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* **2**, 563-572 (2002).
16. Gorgoulis, V. G. et al. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907-913 (2005).
17. Hernando, E. et al. Rb inactivation promotes genomic instability by uncoupling cell cycle progression from mitotic control. *Nature* **430**, 797-802 (2004).
18. Maser, R. S. & DePinho, R. A. Connecting chromosomes, crisis, and cancer. *Science* **297**, 565-569 (2002).
19. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767 (1990).
20. Klein, C. A. Selection and adaptation during metastatic cancer progression. *Nature* **501**, 365-372 (2013).
21. Baker, A. M., Bird, D., Lang, G., Cox, T. R. & Erler, J. T. Lysyl oxidase enzymatic function increases stiffness to drive colorectal cancer progression through FAK. *Oncogene* **32**, 1863-1868 (2013).
22. Barker, H. E., Bird, D., Lang, G. & Erler, J. T. Tumor-secreted LOXL2 activates fibroblasts through FAK signaling. *Mol Cancer Res* **11**, 1425-1436 (2013).
23. Brognard, J. & Hunter, T. Protein kinase signaling networks in cancer. *Curr Opin Genet Dev* **21**, 4-11 (2011).
24. Cox, T. R. et al. LOX-mediated collagen crosslinking is responsible for fibrosis-enhanced metastasis. *Cancer Res* **73**, 1721-1732 (2013).
25. Criscuoli, M. L., Nguyen, M. & Eliceiri, B. P. Tumor metastasis but not tumor growth is dependent on Src-mediated vascular permeability. *Blood* **105**, 1508-1514 (2005).
26. Engelman, J. A. et al. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316**, 1039-1043 (2007).
27. Jorgensen, C. et al. Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells. *Science* **326**, 1502-1509 (2009).
28. Lee, M. J. et al. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780-794 (2012).
29. Regan Anderson, T. M. et al. Breast tumor kinase (Brk/PTK6) is a mediator of hypoxia-associated breast cancer progression. *Cancer Res* **73**, 5810-5820 (2013).
30. Davis, M. I. et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1046-1051 (2011).
31. Fedorov, O., Muller, S. & Knapp, S. The (un)targeted cancer kinome. *Nat Chem Biol* **6**, 166-169 (2010).

32. Janne, P. A., Gray, N. & Settleman, J. Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. *Nat Rev Drug Discov* **8**, 709-723 (2009).
33. Karaman, M. W. et al. A quantitative analysis of kinase inhibitor selectivity. *Nat Biotechnol* **26**, 127-132 (2008).
34. Lin, J. et al. A multidimensional analysis of genes mutated in breast and colorectal cancers. *Genome Res* **17**, 1304-1318 (2007).
35. Wood, L. D. et al. The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108-1113 (2007).
36. Linding, R. et al. Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415-1426 (2007).
37. Pawson, T. & Kofler, M. Kinome signaling through regulated protein-protein interactions in normal and cancer cells. *Curr Opin Cell Biol* **21**, 147-153 (2009).
38. Leibovitz, A. et al. Classification of human colorectal adenocarcinoma cell lines. *Cancer Res* **36**, 4562-4569 (1976).
39. Ong, S. E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).
40. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P. & Jorgensen, T. J. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* **4**, 873-886 (2005).
41. Fabbro, D., Cowan-Jacob, S. W., Mobitz, H. & Martiny-Baron, G. Targeting cancer with small-molecular-weight kinase inhibitors. *Methods Mol Biol* **795**, 1-34 (2012).
42. Davis, M. I. et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1046-1051 (2011).
43. Carrato, A. et al. Fluorouracil, leucovorin, and irinotecan plus either sunitinib or placebo in metastatic colorectal cancer: a randomized, phase III trial. *J Clin Oncol* **31**, 1341-1347 (2013).
44. Saltz, L. B. et al. Phase II trial of sunitinib in patients with metastatic colorectal cancer after failure of standard therapy. *J Clin Oncol* **25**, 4793-4799 (2007).
45. Olsen, J. V. & Macek, B. High accuracy mass spectrometry in large-scale analysis of protein phosphorylation. *Methods Mol Biol* **492**, 131-142 (2009).
46. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-1906 (2007).
47. Swaney, D. L., McAlister, G. C. & Coon, J. J. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods* **5**, 959-964 (2008).
48. Vizcaino, J. A. et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* **41**, D1063-9 (2013).

Figure 1

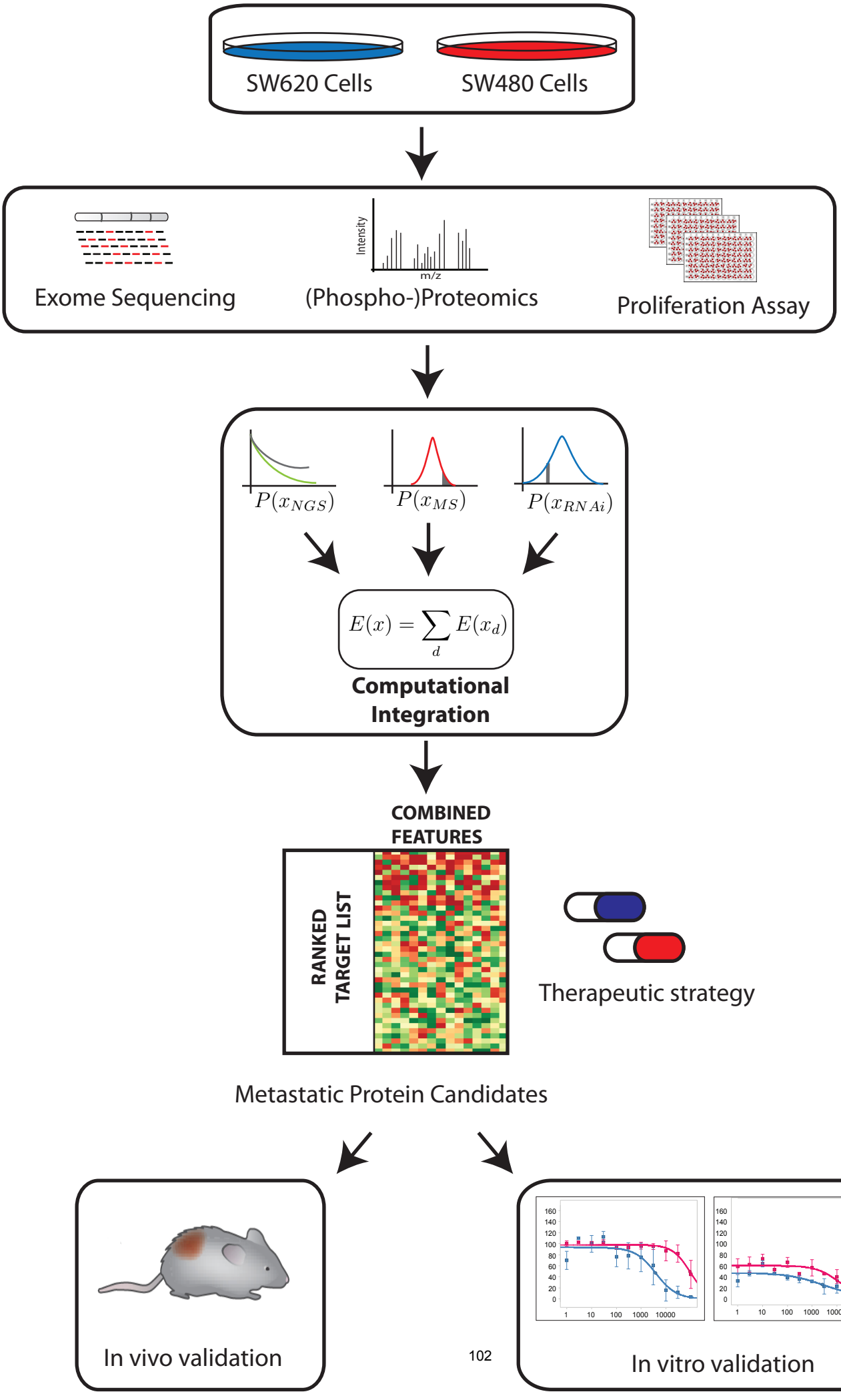
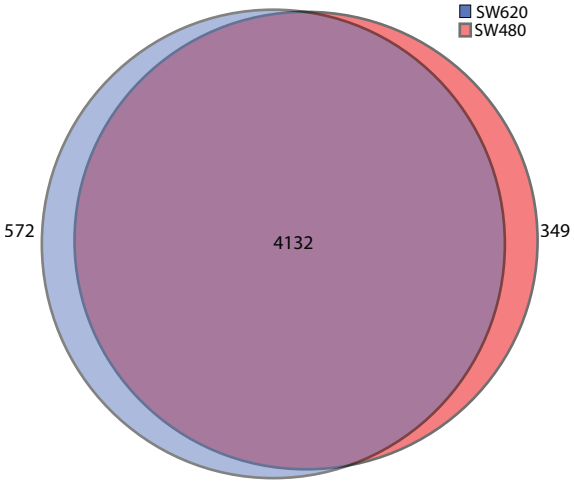

SW620 Cells
SW480 Cells
Exome Sequencing
Intensity
m/z
(Phospho-)Proteomics
Proliferation Assay
$P(x_{NGS})$
$P(x_{MS})$
$P(x_{RNAi})$
$E(x) = \sum_d E(x_d)$
Computational Integration
COMBINED FEATURES
RANKED TARGET LIST
Therapeutic strategy
Metastatic Protein Candidates
In vivo validation
In vitro validation

Figure 2

A)

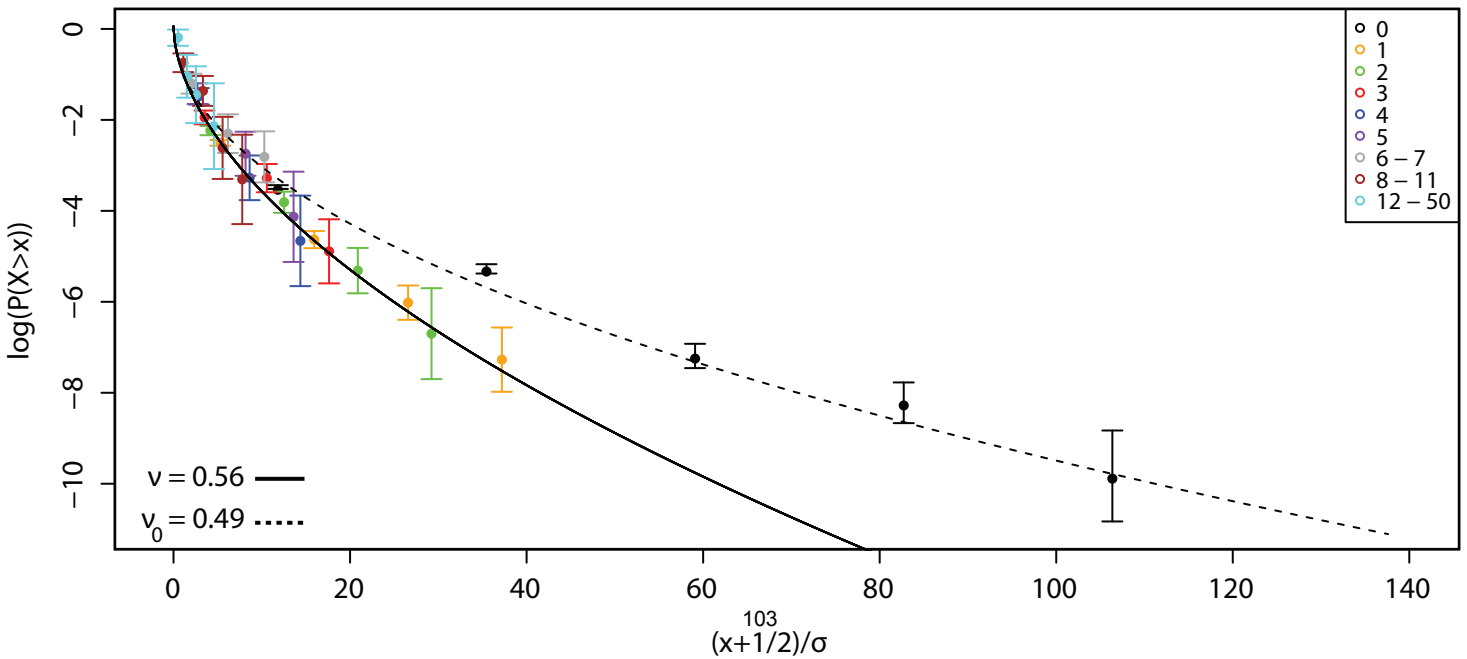
	Nr. Of Mutations	Nr. Of Mutated Proteins	Nr. Of Mutated Kinases
SW480	7240	4481	89
SW620	7680	4704	100

B)

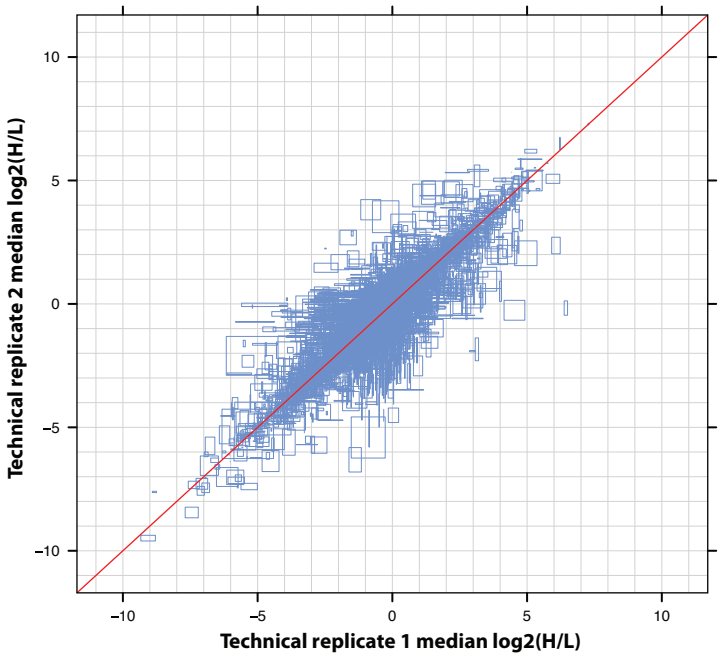


Nr. of Mutated Proteins

C)



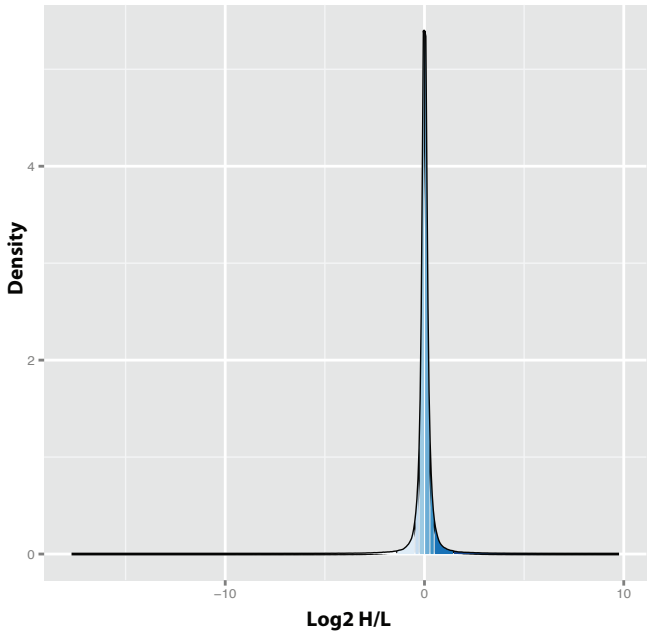
A)



B)

Quantiles:

1%	5%	10%	50%	90%	95%	99%
-1.416762891	-0.479232929	-0.27722375	0.007587979	0.300160955	0.506316838	1.432794152



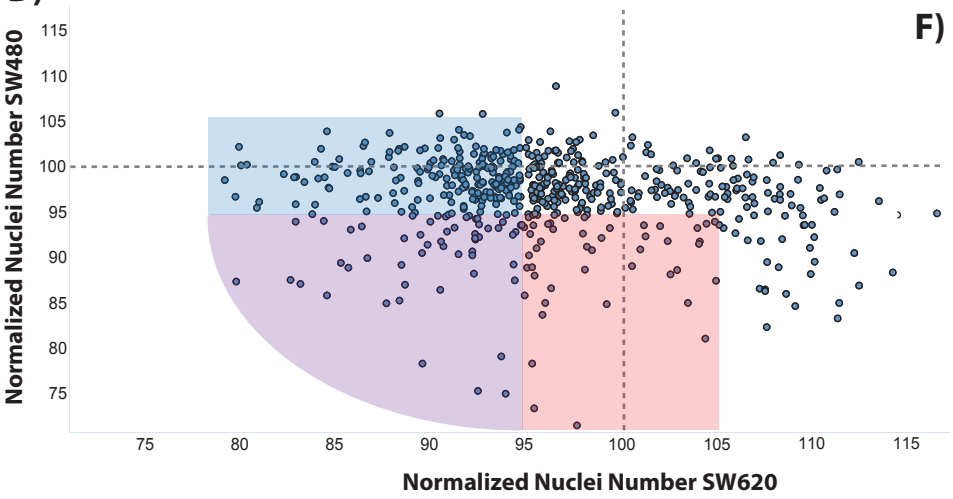
C)

Total Phosphorylation Sites Identified:	28,260
Total Proteins Identified:	9,070
Nr of Kinases Identified:	271

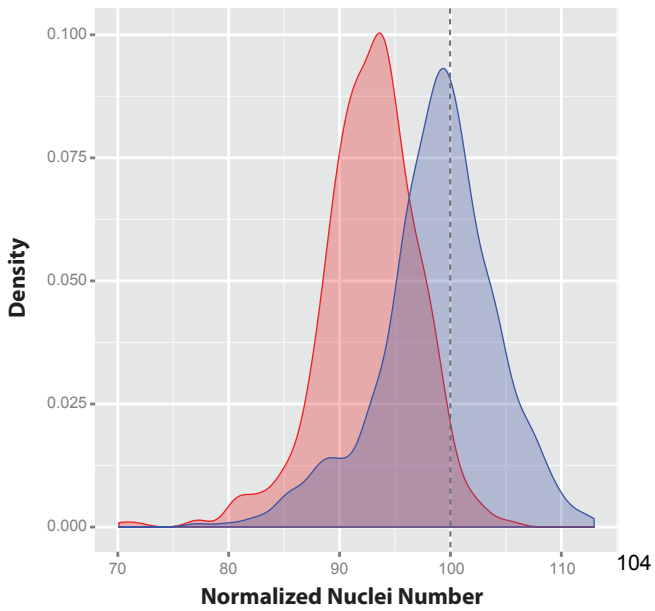
	Up	Down	Steady-State
Phosphorylation Sites	3224	1513	9962
Proteins	358	133	5192

Class I, N=3

D)



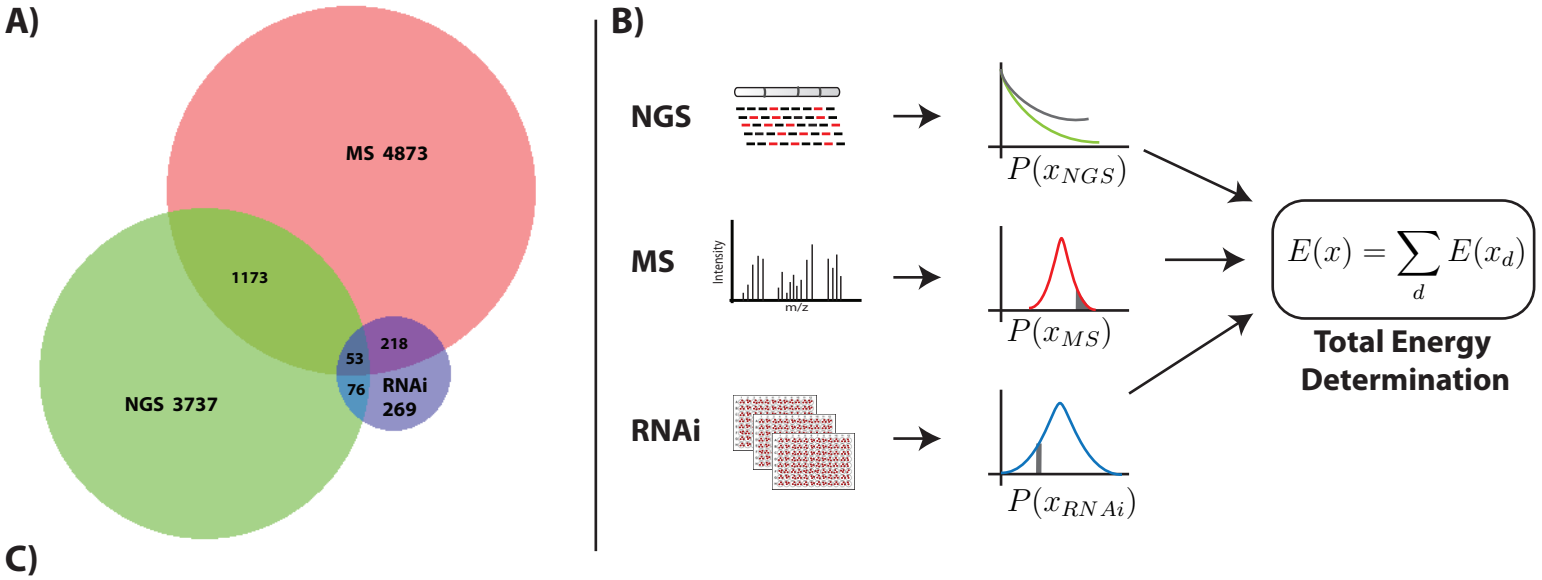
E)



F)

GeneSymbol	SW480	SW620	SW620 / SW480
SIK3	102.4722769	80.22082867	0.782853969
EIF2AK1	100.4153094	80.34075015	0.800084674
FYN	100.5319988	80.63837003	0.802116451
CAMK2B	98.83844068	79.47858467	0.80412625
FLT4	104.1819791	84.87748342	0.81470408
CDC2L2	96.99141221	80.04749332	0.825304957
MYLK4	102.2777388	84.56232836	0.826791141
JUN	99.48577741	82.58405487	0.830109158
LMTK3	100.8348172	84.22326696	0.835259777
DCLK3	99.11421201	83.13609609	0.83879087
FES	99.146463	83.23185164	0.839483821
HUNK	96.40373828	81.29998601	0.843328148
MAP3K13	102.9411101	86.86361429	0.843818512
HIPK3	101.1460128	85.50601074	0.845372036
BRSK1	102.5578123	86.7826343	0.846182581
ERN1	104.0221892	88.17049352	0.847612362
NUAK1	95.75735291	81.17997858	0.847767572
CDC42BPG	98.59243363	83.68310102	0.848778126
FRK	100.3039051	85.16386682	0.849058337
KDR	100.2550644	85.27930771	0.85062344
TAOK1	98.97041334	84.41199968	0.852901355
EEF2K	99.12081234	84.73472634	0.854863114
RAC1	106.1342075	90.79661448	0.855488693
TSSK4	99.62206918	85.96097473	0.862870802
TNNI3K	100.8227396	87.21022085	0.864985629
TNK1	101.9203475	88.17460072	0.865132458
GRK1	96.16220462	83.22087182	0.865421838
MAPK1	102.2830402	88.6661749	0.866870741
ADCK3	99.86996742	86.6331042	0.867459021
GAK	97.91697361	84.98608099	0.867940234
MAP3K10	102.4861885	88.95431465	0.867963927
RPS6KA3	102.0378042	88.61958917	0.868497611
RPS6KB1	101.3278341	88.12922401	0.86974349
MAST4	99.56686955	86.64828833	0.870252211
CAMK1	99.78200641	87.08466526	0.87274919
TNIK	96.09795163	84.24462955	0.876653749

Figure 3



C)

HUGO	ensembl_gene_id	ENSP	energy.total	energy.siRNA	energy.mut	energy.ratio.mean	energy.occ.on.median	siRNA	ratio.mean	occ.on.median	Ncommon	Nonly480	Nonly620	N480	N620
MYO3A	ENSG00000095777	ENSP00000265944	12.5184377	-1.267914698	13.7863524	0	0	0.984436999	NA	NA	2	0	4	2	6
PXK	ENSG00000168297	ENSP00000373222	11.68074491	-1.257883542	11.48007852	1.458549926	0	0.985269053	-1.390882625	NA	0	0	2	0	2
STK36	ENSG00000163482	ENSP00000295709	6.234818041	-0.313214213	6.511974147	0	0.034038843	0.910692539	NA	-0.796122833	3	2	2	5	5
TRIO	ENSG00000038382	ENSP00000339299	5.985000738	-1.405915217	6.471062514	0.291696434	0.23528193	0.960124387	1.075058325	2.143109323	0	0	1	0	1
FYN	ENSG00000010810	ENSP00000346671	5.831757079	4.348756161	-0.333209043	1.566127903	0.08847236	0.802116451	2.03856975	1.672409421	0	0	0	0	0
MYLK3	ENSG00000140795	ENSP00000378288	5.775579072	-0.695483442	6.471062514	0	0	0.921336887	NA	NA	0	0	1	0	1
PRKDC	ENSG00000253729	ENSP00000313420	5.716079357	-0.055470573	6.471062514	-0.225273636	-0.147775654	0.904265132	-0.4029464	-0.366447775	0	0	1	0	1
DCLK1	ENSG00000133083	ENSP00000369223	5.587332765	-1.393405028	-0.333209043	3.311016771	1.541398877	0.956716446	-2.807691333	-4.52008751	0	0	0	0	0
WNK3	ENSG00000196632	ENSP00000364312	5.42165252	-1.049409994	6.471062514	0	0	0.933707323	NA	NA	0	0	1	0	1
PRKG2	ENSG00000138669	ENSP00000378945	5.29364878	-1.177413734	6.471062514	0	0	0.939509558	NA	NA	0	0	1	0	1
FYN	ENSG00000010810	ENSP00000357667	4.806125932	4.348756161	-0.333209043	0.790578813	0	0.802116451	1.397338	NA	0	0	0	0	0
BMPR2	ENSG00000204217	ENSP00000363708	4.548149437	-1.292642033	6.471062514	0	-0.226697392	0.946278419	NA	0.663236732	0	0	1	0	1
SIK3	ENSG00000160584	ENSP00000364449	4.54194408	5.446020065	-0.253231502	0	-0.226690159	0.782853969	NA	-0.057146056	1	0	0	1	1
SIK3	ENSG00000160584	ENSP00000390442	4.228176434	5.446020065	-0.253231502	-0.834250775	-0.096873679	0.782853969	0.255456715	1.158575831	1	0	0	1	1
TNIK	ENSG00000154310	ENSP00000399511	4.123267801	1.042833913	-0.333209043	1.245989585	0.735779846	0.876653749	1.698119	4.200244732	0	0	0	0	0
CAMK2B	ENSG00000058404	ENSP00000326375	3.918654327	4.25186337	-0.333209043	0	0	0.80412625	NA	NA	0	0	0	0	0
ERN1	ENSG00000178607	ENSP00000401445	3.91844541	2.028056764	-0.333209043	2.223597688	0	0.847612362	-2.024685	NA	0	0	0	0	0
NUAK1	ENSG00000074590	ENSP00000261402	3.811445715	2.021758351	-0.333209043	0	0.863751243	0.847767572	NA	-3.27918866	0	0	0	0	0
AURA17	ENSG00000188906	ENSP00000398726	3.722574759	-1.330541758	5.053116518	0	0	0.949186992	NA	NA	1	0	1	1	2
NUAK2	ENSG00000163545	ENSP00000356125	3.539455817	0.920967694	-0.333209043	1.301253376	0.521928962	0.880106036	1.77632	2.930920406	0	0	0	0	0
CDC42BPG	ENSG00000171219	ENSP00000345133	3.500757196	1.981238606	-0.253231502	-0.835339827	0.858291352	0.848778126	0.2385383	4.958262891	1	0	0	1	1
FLT4	ENSG00000037280	ENSP00000261937	3.470468224	3.723699726	-0.253231502	0	0	0.81470408	NA	NA	1	0	0	1	1
EIF2AK1	ENSG00000086232	ENSP00000199389	3.291368538	4.446823013	-0.333209043	-0.822245432	0	0.800084674	0.3197245	NA	0	0	0	0	0
STK35	ENSG00000125834	ENSP00000370891	2.95843729	0.605345329	-0.333209043	2.686301004	0	0.88836859	-2.449177	NA	0	0	0	0	0
CAMK2D	ENSG00000145349	ENSP00000339740	2.918316581	-0.727977016	-0.333209043	2.518435775	0.62389	0.922339566	2.8164385	3.723409875	0	0	0	0	0
CAMKK1	ENSG00000004660	ENSP00000371190	2.781590042	-0.202665991	-0.333209043	-0.588015005	1.362965735	0.90788672	0.56952105	6.565377813	0	0	0	0	0
MYLK4	ENSG00000145949	ENSP00000274643	2.725691163	3.058900205	-0.333209043	0	0	0.826791141	NA	NA	0	0	0	0	0
LMTK3	ENSG00000142235	ENSP00000270238	2.266926686	2.600135728	-0.333209043	0	0	0.835259777	NA	NA	0	0	0	0	0
DCLK3	ENSG00000163673	ENSP00000394484	2.089528569	2.422737611	-0.333209043	0	0	0.83879087	NA	NA	0	0	0	0	0
NEK3	ENSG00000136098	ENSP00000339429	2.063516764	-1.386568088	-0.333209043	2.054202898	0.590036525	0.955508115	-1.9227755	-2.341083009	0	0	0	0	0
FES	ENSG00000182511	ENSP00000410477	2.055900527	2.389109569	-0.333209043	0	0	0.839483821	NA	NA	0	0	0	0	0
CAMK1	ENSG00000134072	ENSP00000256460	2.004825785	1.173799751	-0.333209043	1.164235077	0	0.87274919	1.613832	NA	0	0	0	0	0
HUNK	ENSG00000142149	ENSP00000270112	1.987214086	2.210680448	-0.333209043	-0.234200318	0.144902735	0.843328148	0.7819447	-1.109884713	0	0	0	0	0
EGFR	ENSG00000146648	ENSP00000275493	1.969033019	-1.403251083	-0.333209043	2.016326376	0.671570918	0.959145682	-1.89128875	-2.849136297	0	0	0	0	0
MAP3K13	ENSG00000073803	ENSP00000392223	1.85570125	2.188910293	-0.333209043	0	0	0.843818512	NA	NA	0	0	0	0	0
HIPK3	ENSG00000110422	ENSP00000431710	1.788232143	2.121441185	-0.333209043	0	0	0.845372036	NA	NA	0	0	0	0	0
BRSK1	ENSG00000160469	ENSP00000310649	1.753915358	2.087124401	-0.333209043	0	0	0.846182581	NA	NA	0	0	0	0	0
KDR	ENSG00000128052	ENSP00000263923	1.73108093	1.909506233	-0.178425303	0	0	0.85062344	NA	NA	2	0	0	2	2
TNK1	ENSG00000174292	ENSP00000459799	1.578414413	1.417373481	-0.333209043	0.168927614	0.117190732	0.865132458	1.0067846	1.745366216	0	0	0	0	0

Covered by inhibitors tested

Tested unsuccessfully

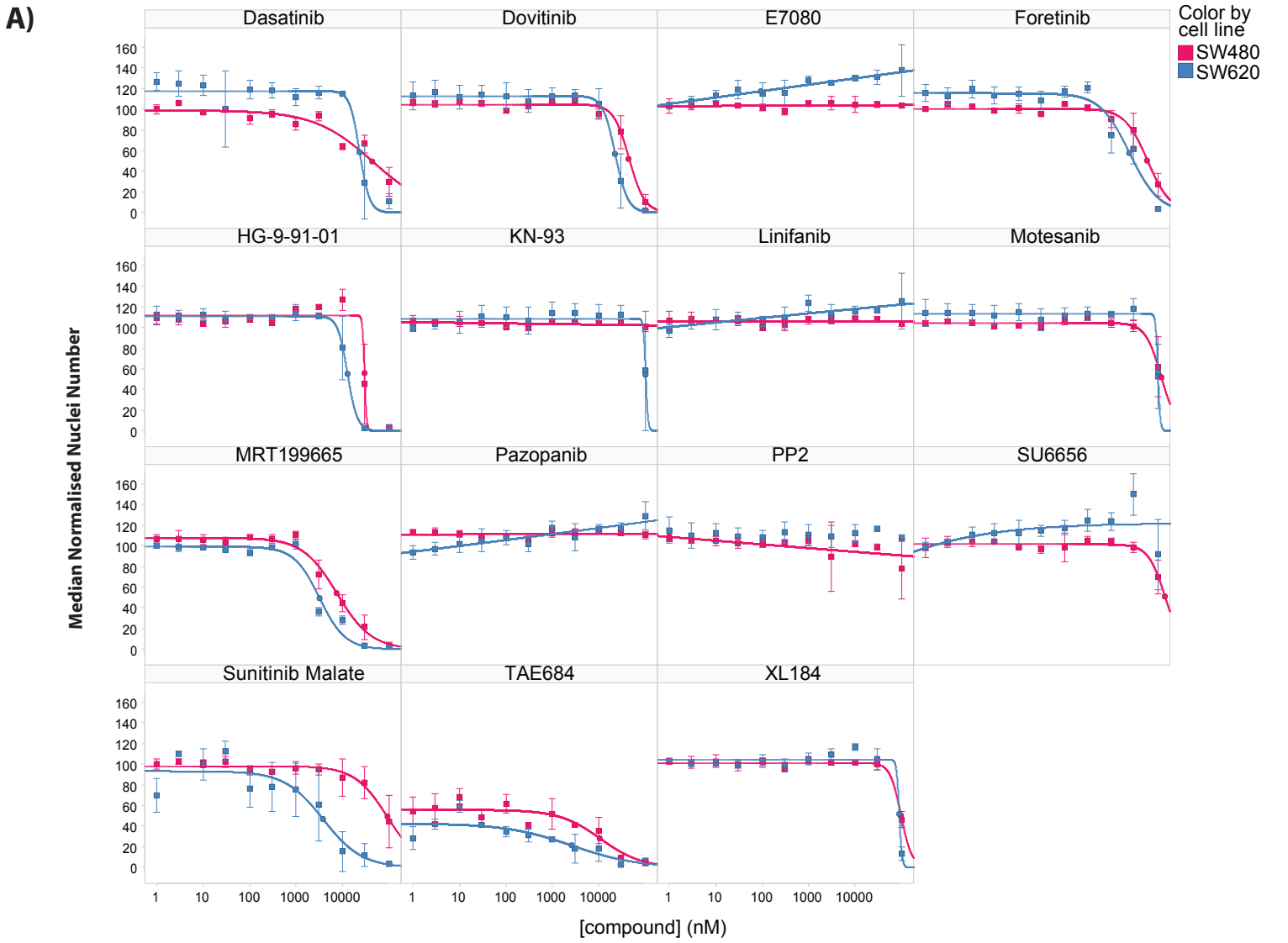
No inhibitors available

Not tested

	Dasatanib	Dovitinib	E7080 (IC50)	Foretinib	HG-9-91-01*	KN-93 (IC-50)	Linifanib	Motesanib	MRT199665*	Pazopanib	PP2 (IC-50)	SU6656 (IC-50)	Sunitinib	TAE684	XL184 (IC-50)
MYO3A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PXK	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
STK36	210	-	-	180	-	-	-	-	-	470	-	-	-	1400	-
TRIO	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FYN	0.79	440	-	88	-	-	-	2800	-	2700	5	170	520	1400	-
MYLK3	-	2	-	3000	-	-	-	-	-	-	-	-	23	-	-
PRKDC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DCLK1	-	-	-	-	-	-	-	-	-	-	-	-	370	4.9	-
WNK3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PRKG2	-	-	-	-	-	-	-	-	-	-	-	-	-	240	-
BMPR2	-	3100	-	-	-	-	7800	-	-	-	-	-	570	3900	-
SIK3	-	-	-	-	5	-	-	-	5	-	-	-	-	-	-
TNIK	2000	24	-	210	-	-	2800	-	-	310	-	-	25	1200	-
CAMK2B	-	-	-	-	-	370	-	-	-	-	-	-	1400	1900	-
ERN1	-	-	-	-	-	-	-	-	-	-	-	-	600	180	-
NUAK1	-	240	-	-	64	-	-	-	2	-	-	-	48	13	-
AURA17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NUAK2	-	130	-	-	-	-	-	-	-	-	-	-	150	1.2	-
CDC42BPG	1200	-	-	180	-	-	8300	-	-	-	-	-	-	9500	-
FLT4	-	580	5.2	1.5	-	-	16	9.7	-	27	-	-	50	170	6
EIF2AK1	-	250	-	980	-	-	-	-	-	-	-	-	-	-	-
STK35	770	-	-	230	-	-	-	5400	-	-	-	-	1300	260	-
CAMK2D	-	-	-	-	-	-	-	-	-	-	-	-	420	570	-
CAMKK1	-	3100	-	550	-	-	-	-	-	-	-	-	420	50	-
MYLK4	-	80	-	-	-	-	-	-	-	-	-	-	15	-	-
LMTK3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DCLK3	-	1300	-	-	-	-	-	-	-	-	-	-	110	14	-
NEK3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FES	-	-	-	110	-	-	-	-	-	1400	-	-	960	4.8	-
CAMK1	-	-	-	5900	90	-	-	-	61	2100	-	-	970	1100	-
HUNK	-	410	-	5400	-	-	-	-	-	-	-	-	500	350	-
EGFR	120	-	-	440	-	-	-	1300	-	-	480	-	860	180	-
MAP3K13	5300	170	-	16	-	-	-	-	-	-	-	-	95	57	-
HIPK3	-	5100	-	43	74	-	2300	-	98	-	-	-	41	6900	-
BRSK1	-	-	-	-	86	-	-	88	-	-	-	-	3500	310	-
KDR	2900	68	-	12	-	-	8.1	26	-	14	-	-	1.5	940	0.035
TNK1	-	-	-	21	-	-	-	-	-	-	-	-	680	1.8	-

* % activity remaining at 1um

Figure 4


A)
Dasatinib
Dovitinib
E7080
Foretinib
HG-9-91-01
KN-93
Linifanib
Motesanib
MRT199665
Pazopanib
PP2
SU6656
Sunitinib Malate
TAE684
XL184
Color by cell line
SW480
SW620
Median Normalised Nuclei Number
[compound] (nM)

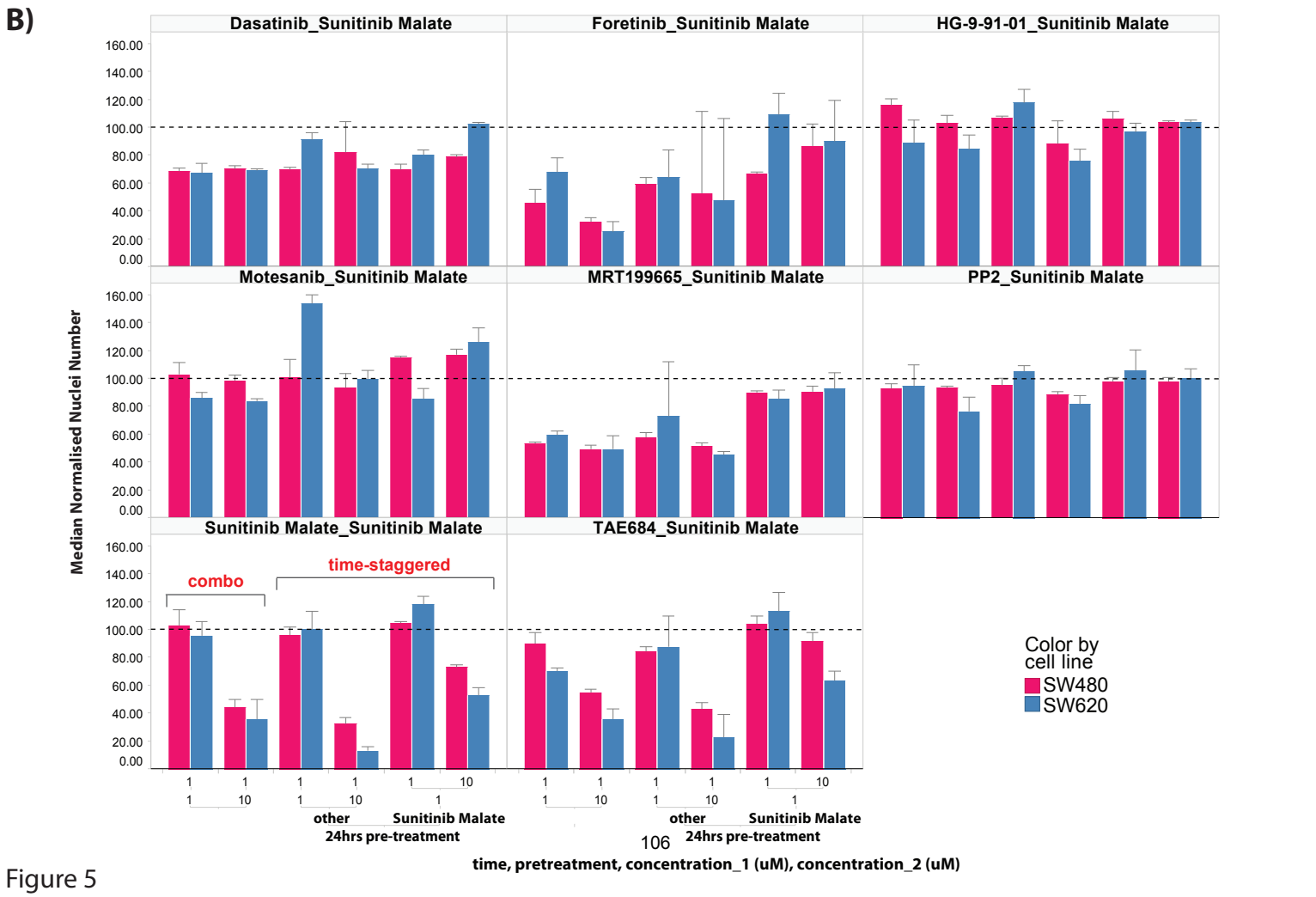

B)
Dasatinib_Sunitinib Malate
Foretinib_Sunitinib Malate
HG-9-91-01_Sunitinib Malate
Motesanib_Sunitinib Malate
MRT199665_Sunitinib Malate
PP2_Sunitinib Malate
Sunitinib Malate_Sunitinib Malate
TAE684_Sunitinib Malate
combo
time-staggered
Median Normalised Nuclei Number
other
Sunitinib Malate
24hrs pre-treatment
time, pretreatment, concentration_1 (uM), concentration_2 (uM)
Color by cell line
SW480
SW620

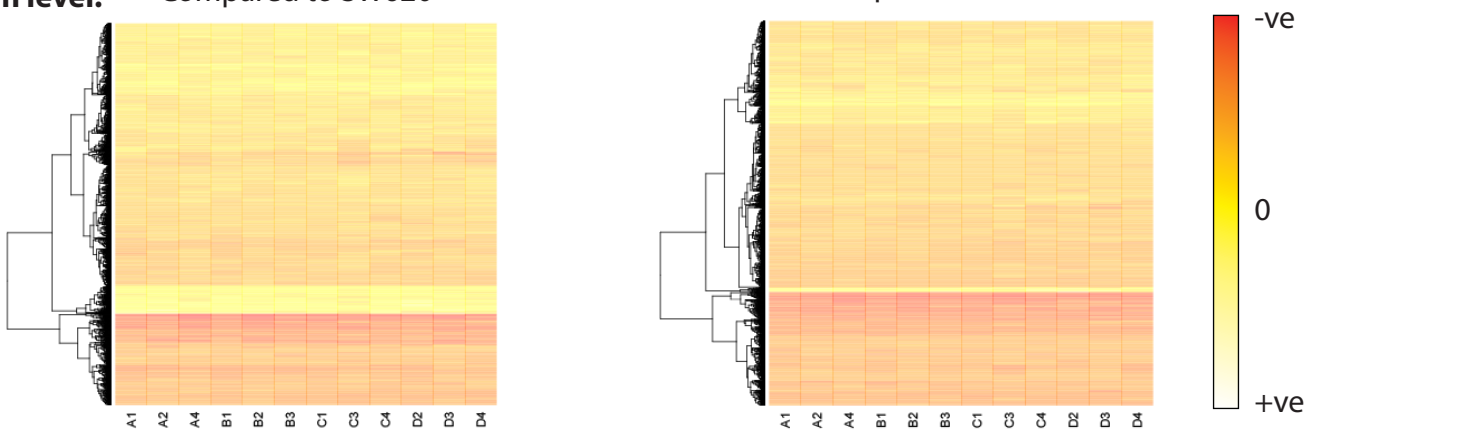
Figure 5

A)

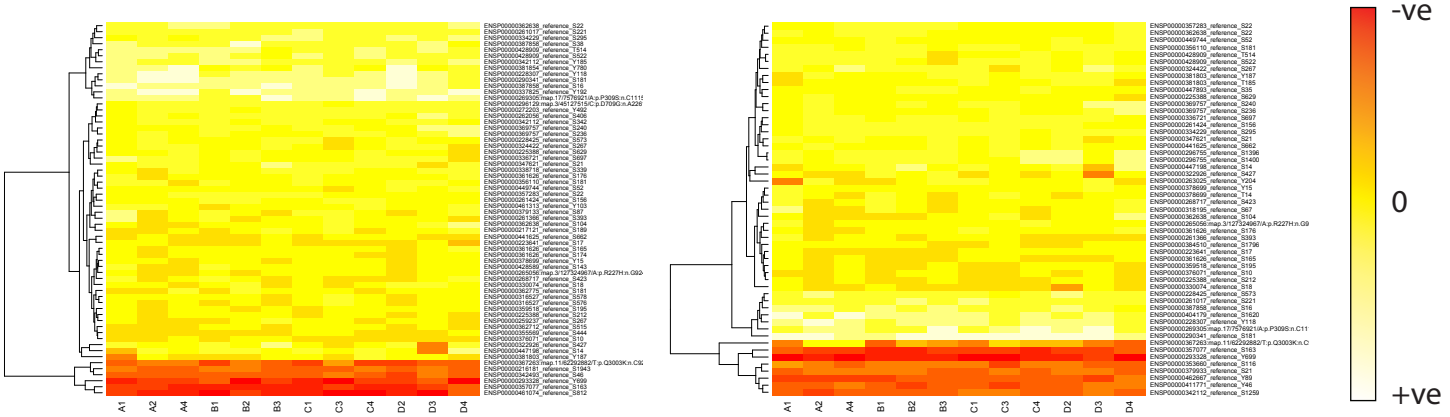
	A1	A2	A4	B1	B2	B3	C1	C3	C4	D2	D3	D4
Nr of Quantifiable Class I Phosphorylation Site IDs:	7492	7585	3585	6923	4909	6219	5215	3237	8413	4179	4735	684
Nr of Quantifiable Protein IDs:	3374	3991	3011	3595	3279	3497	3543	2434	3538	3946	3187	2476
Nr of Non-Synonymous Mutations:	8929	9030	11586	8653	8836	5707	8832	9255	8771	9046	8838	9011
Nr of Mutated Proteins:	5295	5325	6426	5196	5232	3730	5247	5521	5201	5315	5155	5343

Protein level: Compared to SW620 Compared to SW480

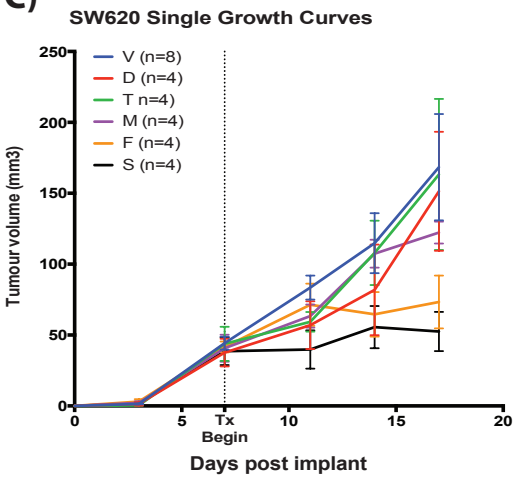
B)

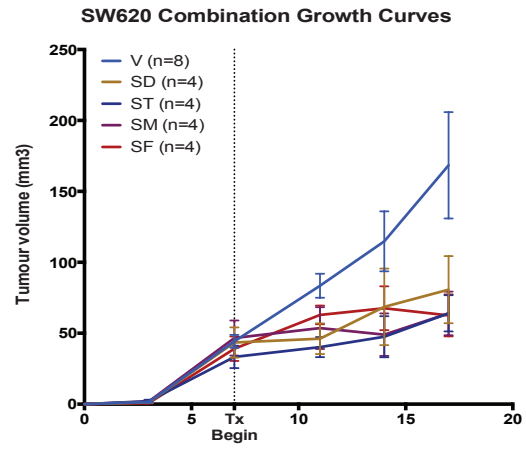


Phospho level:

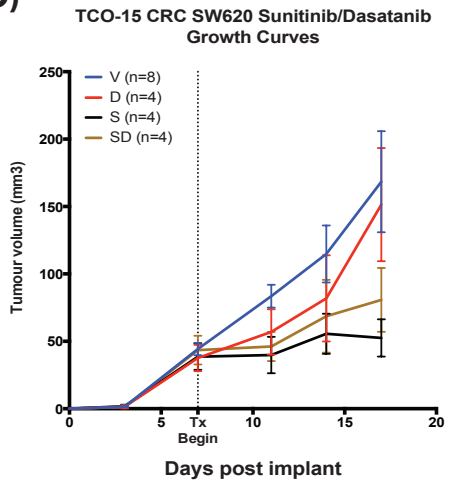


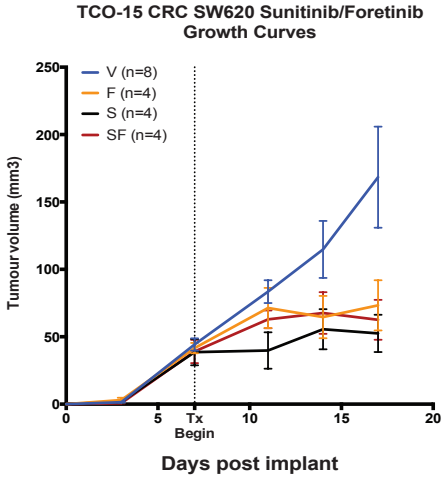
C)

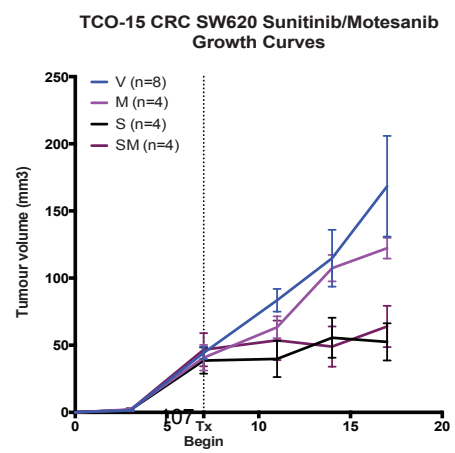




D)







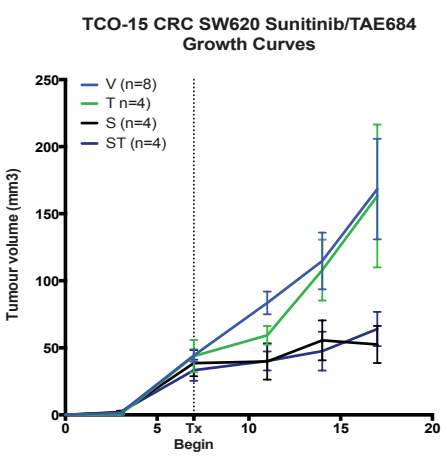


Figure 6

Supplementary Table 1

Sample:	Mutation	Kinase Effect
SW480	ENSP00000166244:map.1/22920150/G:p.Q525R:n.A1646G:c.cAg/cGg:SIFTprediction.tolerated:PolyPhenScore.0.15	hits the kinase protein EphA8 outside its kinase domain
SW480	ENSP00000166244:map.1/22927870/T:p.G936V:n.G2879T:c.gGg/gTg:SIFTprediction.deleterious:PolyPhenScore.0.506	hits the kinase protein EphA8 outside its kinase domain
SW480	ENSP00000216911:map.20/54961463/C:p.I57V:n.A437G:c.Att/Gtt:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein AurA outside its kinase domain
SW480	ENSP00000220751:map.8/90784009/G:p.P228A:n.C996G:c.Cct/Gct:SIFTprediction.deleterious:PolyPhenScore.1	hits the kinase domain of RIPK2
SW480	ENSP00000224764:map.10/88635779/A:p.P2T:n.C552A:c.Cct/Act:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein BMPR1A outside its kinase domain
SW480	ENSP00000226094:map.17/66533655/G:p.L530S:n.T1877C:c.tTg/tCg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein FAM20A outside its kinase domain
SW480	ENSP00000233027:map.3/52797634/C:p.P225A:n.C876G:c.Cca/Gca:SIFTprediction.deleterious:PolyPhenScore.0.784	hits the kinase domain of NEK4
SW480	ENSP00000240361:map.17/56659018/T:p.G1088D:n.G3349A:c.gGt/gAt:SIFTprediction.tolerated:PolyPhenScore.0.033	hits the kinase protein SgK307 outside its kinase domain
SW480	ENSP00000241453:map.13/28624294/A:p.T227M:n.C762T:c.aCg/aTg:SIFTprediction.deleterious:PolyPhenScore.0.803	hits the kinase protein FLT3 outside its kinase domain
SW480	ENSP00000259750:map.6/43230970/C:p.G623A:n.G1951C:c.gGc/gCc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein TTBK1 outside its kinase domain
SW480	ENSP00000260404:map.15/40564576/T:p.P337L:n.C1436T:c.cCg/cTg:SIFTprediction.deleterious:PolyPhenScore.0.018	hits the kinase protein PAK6 outside its kinase domain
SW480	ENSP00000261170:map.12/14829893/C:p.F281L:n.T980G:c.ttT/ttG:SIFTprediction.tolerated:PolyPhenScore.0.02	hits the kinase protein HSER outside its kinase domain
SW480	ENSP00000261937:map.5/180046344/C:p.H890Q:n.C2749G:c.caC/caG:SIFTprediction.deleterious:PolyPhenScore.0.32	hits the kinase domain of FLT4
SW480	ENSP00000262811:map.19/18255359/A:p.G861S:n.G2581A:c.Ggc/Agc:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein MAST3 outside its kinase domain
SW480	ENSP00000263026:map.16/22269867/G:p.Q361R:n.A1556G:c.cAa/cGa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein eEF2K outside its kinase domain
SW480	ENSP00000263791:map.15/40265799/G:p.E556G:n.A1710G:c.gAa/gGa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein GCN2 outside its kinase domain
SW480	ENSP00000263923:map.4/55972974/A:p.Q472H:n.A1712T:c.caA/caT:SIFTprediction.tolerated:PolyPhenScore.0.012	hits the kinase protein KDR outside its kinase domain
SW480	ENSP00000263923:map.4/55979558/T:p.V297I:n.G1185A:c.Gta/Ata:SIFTprediction.tolerated:PolyPhenScore.0.999	hits the kinase protein KDR outside its kinase domain
SW480	ENSP00000263955:map.2/197002262/T:p.I343N:n.T1315A:c.aTc/aAc:SIFTprediction.deleterious:PolyPhenScore.0.728	hits the kinase protein DRAK2 outside its kinase domain
SW480	ENSP00000264316:map.4/48115264/T:p.R45H:n.G220A:c.cGt/cAt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein TXK outside its kinase domain
SW480	ENSP00000265944:map.10/26446312/A:p.S956N:n.G3033A:c.aGt/aAt:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein MYO3A outside its kinase domain
SW480	ENSP00000265944:map.10/26463043/T:p.T1284S:n.A4016T:c.Act/Tct:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein MYO3A outside its kinase domain
SW480	ENSP00000270162:map.21/44837555/A:p.A615V:n.C1977T:c.gCc/gTc:SIFTprediction.deleterious:PolyPhenScore.0.031	hits the kinase protein SIK outside its kinase domain
SW480	ENSP00000270162:map.21/44846016/T:p.G15S:n.G176A:c.Ggt/Agt:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein SIK outside its kinase domain
SW480	ENSP00000275815:map.7/143088867/C:p.M900V:n.A2785G:c.Atg/Gtg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein EphA1 outside its kinase domain
SW480	ENSP00000275815:map.7/143097100/G:p.V160A:n.T566C:c.gTg/gCg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein EphA1 outside its kinase domain
SW480	ENSP00000278616:map.11/108183167/G:p.N1983S:n.A6333G:c.aAt/aGt:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein ATM outside its kinase domain
SW480	ENSP00000283109:map.5/96503523/T:p.G349R:n.G1114A:c.Ggg/Agg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein RIOK2 outside its kinase domain
SW480	ENSP00000283109:map.5/96513471/C:p.S96C:n.C356G:c.tCt/tGt:SIFTprediction.deleterious:PolyPhenScore.0.987	hits the kinase domain of RIOK2
SW480	ENSP00000288135:map.4/55593464/C:p.M541L:n.A1718C:c.Atg/Ctg:SIFTprediction.tolerated:PolyPhenScore.0.008	hits the kinase protein KIT outside its kinase domain
SW480	ENSP00000291281:map.19/47177913/G:p.V835A:n.T3261C:c.gTg/gCg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein PKD2 outside its kinase domain
SW480	ENSP00000291823:map.19/40886993/T:p.R302Q:n.G1190A:c.cGg/cAg:SIFTprediction.tolerated:PolyPhenScore.0.007	hits the kinase domain of HIPK4
SW480	ENSP00000295709:map.2/219544388/G:p.K295R:n.A1163G:c.aAg/aGg:SIFTprediction.tolerated:PolyPhenScore.0.016	hits the kinase protein Fused outside its kinase domain
SW480	ENSP00000295709:map.2/219553468/T:p.R477W:n.C1708T:c.Cgg/Tgg:SIFTprediction.deleterious:PolyPhenScore.0.975	hits the kinase protein Fused outside its kinase domain
SW480	ENSP00000295709:map.2/219555262/A:p.R583Q:n.G2027A:c.cGg/cAg:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein Fused outside its kinase domain
SW480	ENSP00000295709:map.2/219559249/A:p.S801N:n.G2681A:c.aGt/aAt:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein Fused outside its kinase domain
SW480	ENSP00000295709:map.2/219563602/A:p.R1112Q:n.G3614A:c.cGg/cAg:SIFTprediction.tolerated:PolyPhenScore.0.004	hits the kinase protein Fused outside its kinase domain
SW480	ENSP00000297293:map.7/97822115/A:p.L780M:n.T2631A:c.Ttg/Atg:SIFTprediction.tolerated:PolyPhenScore.0.34	hits the kinase protein LMR2 outside its kinase domain
SW480	ENSP00000298910:map.12/40619082/A:p.R50H:n.G207A:c.cGc/cAc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein LRRK2 outside its kinase domain
SW480	ENSP00000301178:map.19/41743861/G:p.N266D:n.A986G:c.Aac/Gac:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AXL outside its kinase domain
SW480	ENSP00000301831:map.3/41756965/T:p.V851I:n.G3014A:c.Gta/Ata:SIFTprediction.tolerated:PolyPhenScore.0.01	hits the kinase protein ULK4 outside its kinase domain
SW480	ENSP00000301831:map.3/41756986/T:p.L844M:n.T2993A:c.Ttg/Atg:SIFTprediction.tolerated:PolyPhenScore.0.004	hits the kinase protein ULK4 outside its kinase domain
SW480	ENSP00000301831:map.3/41831203/T:p.A715T:n.G2606A:c.Gct/Act:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein ULK4 outside its kinase domain
SW480	ENSP00000301831:map.3/41841716/C:p.S640A:n.T2381G:c.Tcc/Gcc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein ULK4 outside its kinase domain
SW480	ENSP00000301831:map.3/41877414/C:p.K569R:n.A2169G:c.aAa/aGa:SIFTprediction.tolerated:PolyPhenScore.0.284	hits the kinase protein ULK4 outside its kinase domain
SW480	ENSP00000301831:map.3/41952852/C:p.S348G:n.A1505G:c.Agt/Ggt:SIFTprediction.deleterious:PolyPhenScore.0.734	hits the kinase protein ULK4 outside its kinase domain
SW480	ENSP00000306678:map.11/113266821/A:p.A239T:n.G809A:c.Gcg/Acg:SIFTprediction.tolerated:PolyPhenScore.0.324	hits the kinase domain of SgK288
SW480	ENSP00000306678:map.11/113270015/C:p.G442R:n.G1418C:c.Ggc/Cgc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein SgK288 outside its kinase domain
SW480	ENSP00000306678:map.11/113270828/A:p.E713K:n.G2231A:c.Gag/Aag:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein SgK288 outside its kinase domain
SW480	ENSP00000307235:map.2/88874891/A:p.A704S:n.G2412T:c.Gct/Tct:SIFTprediction.tolerated:PolyPhenScore.0.005	hits the kinase domain of PEK
SW480	ENSP00000307235:map.2/88895123/C:p.Q166R:n.A799G:c.cAa/cGa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein PEK outside its kinase domain
SW480	ENSP00000308413:map.11/67202156/T:p.A420V:n.C1304T:c.gCc/gTc:SIFTprediction.tolerated:PolyPhenScore.0.01	hits the kinase protein p70S6Kb outside its kinase domain
SW480	ENSP00000309230:map.15/77450964/T:p.R1071K:n.G3491A:c.aGg/aAg:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein SgK269 outside its kinase domain
SW480	ENSP00000310722:map.8/57026229/A:p.A105S:n.G313T:c.Gct/Tct:SIFTprediction.deleterious:PolyPhenScore.0.999	hits the kinase domain of MOS
SW480	ENSP00000317985:map.2/11359120/T:p.T431N:n.C1741A:c.aCt/aAt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein ROCK2 outside its kinase domain
SW480	ENSP00000319192:map.7/43664280/G:p.K362E:n.A1263G:c.Aag/Gag:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein DRAK1 outside its kinase domain
SW480	ENSP00000324560:map.12/132403161/G:p.T816A:n.A2797G:c.Act/Gct:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein ULK1 outside its kinase domain
SW480	ENSP00000332454:map.21/43161357/C:p.M666V:n.A2061G:c.Atg/Gtg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein ANKRD3 outside its kinase domain
SW480	ENSP00000335347:map.14/103934488/C:p.F433S:n.T1298C:c.tTt/tCt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MARK3 outside its kinase domain
SW480	ENSP00000337451:map.3/89521664/A:p.R914H:n.G2966A:c.cGc/cAc:SIFTprediction.tolerated:PolyPhenScore.0.017	hits the kinase protein EphA3 outside its kinase domain
SW480	ENSP00000337451:map.3/89521693/C:p.W924R:n.T2995C:c.Tgg/Cgg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein EphA3 outside its kinase domain
SW480	ENSP00000342105:map.12/68051420/A:p.H245N:n.C1146A:c.Cac/Aac:SIFTprediction.tolerated:PolyPhenScore.0.017	hits the kinase domain of DYRK2
SW480	ENSP00000345083:map.17/21202191/A:p.P40T:n.C367A:c.Ccc/Acc:SIFTprediction.deleterious:PolyPhenScore.0.904	hits the kinase protein MAP2K3 outside its kinase domain
SW480	ENSP00000345083:map.17/21202237/C:p.R55T:n.G413C:c.aGa/aCa:SIFTprediction.tolerated:PolyPhenScore.0.032	hits the kinase protein MAP2K3 outside its kinase domain
SW480	ENSP00000345083:map.17/21203893/C:p.S68P:n.T451C:c.Tca/Cca:SIFTprediction.tolerated:PolyPhenScore.0.065	hits the kinase domain of MAP2K3
SW480	ENSP00000345083:map.17/21203941/A:p.A84T:n.G499A:c.Gcc/Acc:SIFTprediction.tolerated:PolyPhenScore.0.371	hits the kinase domain of MAP2K3
SW480	ENSP00000345083:map.17/21204187/T:p.R94L:n.G530T:c.cGg/cTg:SIFTprediction.deleterious:PolyPhenScore.0.662	hits the kinase domain of MAP2K3
SW480	ENSP00000345083:map.17/21204192/T:p.R96W:n.C535T:c.Cgg/Tgg:SIFTprediction.deleterious:PolyPhenScore.0.996	hits the kinase domain of MAP2K3
SW480	ENSP00000345083:map.17/21207834/T:p.T222M:n.C914T:c.aCg/aTg:SIFTprediction.deleterious:PolyPhenScore.1	hits the kinase domain of MAP2K3
SW480	ENSP00000345083:map.17/21215557/A:p.R293H:n.G1127A:c.cGt/cAt:SIFTprediction.tolerated:PolyPhenScore.0.874	hits the kinase domain of MAP2K3
SW480	ENSP00000345083:map.17/21217513/A:p.V339M:n.G1264A:c.Gtg/Atg:SIFTprediction.deleterious:PolyPhenScore.0.807	hits the kinase protein MAP2K3 outside its kinase domain
SW480	ENSP00000345133:map.11/64597506/C:p.Q1135R:n.A3404G:c.cAg/cGg:SIFTprediction.tolerated:PolyPhenScore.0.008	hits the kinase protein DMPK2 outside its kinase domain
SW480	ENSP00000345629:map.X/19482476/T:p.A192T:n.G574A:c.Gct/Act:SIFTprediction.tolerated:PolyPhenScore.0.408	hits the kinase protein MAP3K7 outside its kinase domain
SW480	ENSP00000348132:map.7/23811795/G:p.N621K:n.T1982G:c.aaT/aaG:SIFTprediction.deleterious:PolyPhenScore.0.713	hits the kinase protein SgK396 outside its kinase domain
SW480	ENSP00000350195:map.1/27687466/T:p.N622K:n.C2135A:c.aaC/aaA:SIFTprediction.tolerated:PolyPhenScore.0.013	hits the kinase protein MAP3K6 outside its kinase domain
SW480	ENSP00000350195:map.1/27688633/A:p.T455I:n.C1633T:c.aCc/aTc:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein MAP3K6 outside its kinase domain
SW480	ENSP00000353452:map.3/123451773/C:p.L496V:n.C1768G:c.Ctg/Gtg:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein smMLCK outside its kinase domain
SW480	ENSP00000353452:map.3/123457893/A:p.P147S:n.C721T:c.Cca/Tca:SIFTprediction.tolerated:PolyPhenScore.0.011	hits the kinase protein smMLCK outside its kinase domain
SW480	ENSP00000354006:map.9/77397374/T:p.S1038Y:n.C3351A:c.tCc/tAc:SIFTprediction.deleterious:PolyPhenScore.0.619	hits the kinase protein ChaK2 outside its kinase domain
SW480	ENSP00000354671:map.1/46476587/G:p.D388E:n.T1447G:c.gaT/gaG:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MAST2 outside its kinase domain
SW480	ENSP00000354671:map.1/46493460/G:p.I659M:n.T2260G:c.aT/atG:SIFTprediction.tolerated:PolyPhenScore.0.491	hits the kinase domain of MAST2
SW480	ENSP00000354877:map.17/19713740/T:p.V370M:n.G1627A:c.Gtg/Atg:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein ULK2 outside its kinase domain
SW480	ENSP00000354991:map.18/56149099/C:p.I2157V:n.A6683G:c.Ata/Gta:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AlphaK2 outside its kinase domain
SW480	ENSP00000354991:map.18/56202768/A:p.A1551S:n.G4865T:c.Gct/Tct:SIFTprediction.deleterious:PolyPhenScore.0.258	hits the kinase protein AlphaK2 outside its kinase domain
SW480	ENSP00000354991:map.18/56203074/A:p.P1449S:n.C4559T:c.Ccg/Tcg:SIFTprediction.deleterious:PolyPhenScore.0.904	hits the kinase protein AlphaK2 outside its kinase domain
SW480	ENSP00000354991:map.18/56204671/T:p.N916K:n.T2962A:c.aaT/aaA:SIFTprediction.deleterious:PolyPhenScore.0.731	hits the kinase protein AlphaK2 outside its kinase domain
SW480	ENSP00000354991:map.18/56204747/A:p.T891I:n.C2886T:c.aCc/aTc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AlphaK2 outside its kinase domain
SW480	ENSP00000354991:map.18/56204932/G:p.K829N:n.A2701C:c.aaA/aaC:SIFTprediction.deleterious:PolyPhenScore.0.824	hits the kinase protein AlphaK2 outside its kinase domain
SW480	ENSP00000354991:map.18/56204945/G:p.R825T:n.G2688C:c.aGa/aCa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AlphaK2 outside its kinase domain
SW480	ENSP00000354991:map.18/56204991/T:p.G810S:n.G2642A:c.Ggt/Agt:SIFTprediction.deleterious:PolyPhenScore.0.951	hits the kinase protein AlphaK2 outside its kinase domain
SW480	ENSP00000354991:map.18/56205262/C:p.H719Q:n.T2371G:c.caT/caG:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AlphaK2 outside its kinase domain
SW480	ENSP00000354991:map.18/56247600/A:p.R136S:n.G622T:c.agG/agT:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AlphaK2 outside its kinase domain
SW480	ENSP00000355304:map.14/102695693/C:p.Q398R:n.A1425G:c.cAg/cGg:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein MOK outside its kinase domain
SW480	ENSP00000355583:map.1/233497978/T:p.K497N:n.G1752T:c.aaG/aaT:SIFTprediction.deleterious:PolyPhenScore.0.972	hits the kinase protein MLK4 outside its kinase domain
SW480	ENSP00000355966:map.1/211840498/C:p.N354S:n.A1200G:c.aAc/aGc:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein NEK2 outside its kinase domain
SW480	ENSP00000356087:map.1/206669465/T:p.P713L:n.C2511T:c.cCt/cTt:SIFTprediction.tolerated:PolyPhenScore.0.008	hits the kinase protein IKKe outside its kinase domain
SW480	ENSP00000356130:map.1/205130413/G:p.C641R:n.T1952C:c.Tgt/Cgt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein Dusty outside its kinase domain
SW480	ENSP00000356530:map.1/182551337/C:p.D541E:n.T1877G:c.gaT/gaG:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase domain of RNAseL
SW480	ENSP00000356530:map.1/182554557/T:p.R462Q:n.G1639A:c.cGa/cAa:SIFTprediction.deleterious:PolyPhenScore.0.85	hits the kinase domain of RNAseL
SW480	ENSP00000356745:map.1/169823718/C:p.Q567R:n.A1915G:c.cAa/cGa:SIFTprediction.tolerated:PolyPhenScore.0.116	hits the kinase protein SCYL3 outside its kinase domain
SW480	ENSP00000357494:map.6/117622233/T:p.D2213N:n.G6836A:c.Gac/Aac:SIFTprediction.tolerated:PolyPhenScore.0.009	hits the kinase domain of ROS
SW480	ENSP00000357494:map.6/117681543/A:p.P1136L:n.C3606T:c.cCa/cTa:SIFTprediction.deleterious:PolyPhenScore.0.995	hits the kinase protein ROS outside its kinase domain
SW480	ENSP00000357494:map.6/117687244/T:p.L936Q:n.T3006A:c.cTg/cAg:SIFTprediction.deleterious:PolyPhenScore.0.999	hits the kinase protein ROS outside its kinase domain
SW480	ENSP00000357615:map.6/116325142/T:p.G122R:n.G811A:c.Gga/Aga:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein FRK outside its kinase domain
SW480	ENSP00000359424:map.10/101977883/T:p.V268I:n.G857A:c.Gta/Ata:SIFTprediction.tolerated:PolyPhenScore.0.005	hits the kinase domain of IKKa
SW480	ENSP00000361025:map.9/136270538/G:p.L679R:n.T2143G:c.cTg/cGg:SIFTprediction.deleterious:PolyPhenScore.0.912	hits the kinase protein SgK071 outside its kinase domain
SW480	ENSP00000362139:map.1/38185723/T:p.R807Q:n.G2420A:c.cGg/cAg:SIFTprediction.deleterious:PolyPhenScore.0.792	hits the kinase domain of EphA10
SW480	ENSP00000362139:map.1/38188740/T:p.V645I:n.G1933A:c.Gtc/Atc:SIFTprediction.tolerated:PolyPhenScore.0.005	hits the kinase domain of EphA10

SW480	ENSP00000362139:map.1/38188787/G:p.L629P:n.T1886C:c.cTg/cCg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein EphA10 outside its kinase domain
SW480	ENSP00000362139:map.1/38227086/T:p.F281I:n.T841A:c.Ttc/Atc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein EphA10 outside its kinase domain
SW480	ENSP00000362702:map.9/127088697/T:p.S199L:n.C811T:c.tCa/tTa:SIFTprediction.deleterious:PolyPhenScore.0.089	hits the kinase domain of NEK6
SW480	ENSP00000364204:map.1/20977000/C:p.N521T:n.A1656C:c.aAt/aCt:SIFTprediction.deleterious:PolyPhenScore.0.007	hits the kinase protein PINK1 outside its kinase domain
SW480	ENSP00000364361:map.2/174128513/T:p.S531L:n.C1670T:c.tCg/tTg:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein ZAK outside its kinase domain
SW480	ENSP00000364860:map.9/94486321/T:p.V819I:n.G2654A:c.Gtc/Atc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein ROR2 outside its kinase domain
SW480	ENSP00000364860:map.9/94495608/C:p.T245A:n.A932G:c.Aca/Gca:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein ROR2 outside its kinase domain
SW480	ENSP00000366488:map.9/71628207/C:p.H268D:n.C833G:c.Cat/Gat:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase domain of PKACg
SW480	ENSP00000369126:map.13/37679268/T:p.D42E:n.C536A:c.gaC/gaA:SIFTprediction.tolerated:PolyPhenScore.0.004	hits the kinase domain of CK1a2
SW480	ENSP00000369375:map.9/27183463/C:p.Q346P:n.A1479C:c.cAg/cCg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein TIE2 outside its kinase domain
SW480	ENSP00000372035:map.13/21562832/T:p.G363S:n.G1493A:c.Ggc/Agc:SIFTprediction.tolerated:PolyPhenScore.0.005	hits the kinase protein LATS2 outside its kinase domain
SW480	ENSP00000372035:map.13/21562948/A:p.A324V:n.C1377T:c.gCg/gTg:SIFTprediction.tolerated:PolyPhenScore.0.031	hits the kinase protein LATS2 outside its kinase domain
SW480	ENSP00000373600:map.15/101606889/A:p.G1938D:n.G6172A:c.gGc/gAc:SIFTprediction.tolerated:PolyPhenScore.0.013	hits the kinase protein LRRK1 outside its kinase domain
SW480	ENSP00000381129:map.4/2990499/T:p.R65L:n.G537T:c.cGt/cTt:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein GPRK4 outside its kinase domain
SW480	ENSP00000381129:map.4/3006043/T:p.A142V:n.C768T:c.gCc/gTc:SIFTprediction.tolerated:PolyPhenScore.0.004	hits the kinase protein GPRK4 outside its kinase domain
SW480	ENSP00000381129:map.4/3015553/A:p.V247I:n.G1082A:c.Gta/Ata:SIFTprediction.tolerated:PolyPhenScore.0.669	hits the kinase domain of GPRK4
SW480	ENSP00000382423:map.5/56177443/A:p.D806N:n.G2416A:c.Gat/Aat:SIFTprediction.tolerated:PolyPhenScore.0.111	hits the kinase protein MAP3K1 outside its kinase domain
SW480	ENSP00000382423:map.5/56177443/A:p.V906I:n.G2716A:c.Gtc/Atc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MAP3K1 outside its kinase domain
SW480	ENSP00000382544:map.22/19119751/T:p.T280M:n.C1431T:c.aCg/aTg:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein TSSK2 outside its kinase domain
SW480	ENSP00000383234:map.18/48190440/A:p.V38M:n.G1112A:c.Gtg/Atg:SIFTprediction.tolerated:PolyPhenScore.0.357	hits the kinase domain of ERK4
SW480	ENSP00000384442:map.1/1650787/C:p.H112R:n.A415G:c.cAt/cGt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein CDK11b outside its kinase domain
SW480	ENSP00000384442:map.1/1650797/G:p.C109R:n.T405C:c.Tgt/Cgt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein CDK11b outside its kinase domain
SW480	ENSP00000384442:map.1/1650832/G:p.V97A:n.T370C:c.gTt/gCt:SIFTprediction.tolerated:PolyPhenScore.0.013	hits the kinase protein CDK11b outside its kinase domain
SW480	ENSP00000384442:map.1/1650845/A:p.R93W:n.C357T:c.Cgg/Tgg:SIFTprediction.deleterious:PolyPhenScore.0.999	hits the kinase protein CDK11b outside its kinase domain
SW480	ENSP00000386135:map.9/90322023/A:p.S1346N:n.G4372A:c.aGt/aAt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein DAPK1 outside its kinase domain
SW480	ENSP00000386213:map.2/171260787/A:p.V770I:n.G2451A:c.Gta/Ata:SIFTprediction.tolerated:PolyPhenScore.0.004	hits the kinase protein MYO3B outside its kinase domain
SW480	ENSP00000386213:map.2/171356274/A:p.R1082K:n.G3388A:c.aGg/aAg:SIFTprediction.tolerated:PolyPhenScore.0.005	hits the kinase protein MYO3B outside its kinase domain
SW480	ENSP00000386456:map.2/69741854/G:p.K509Q:n.A1902C:c.Aaa/Caa:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein AAK1 outside its kinase domain
SW480	ENSP00000389015:map.19/56047448/G:p.C72R:n.T252C:c.Tgc/Cgc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase domain of SgK069
SW480	ENSP00000391295:map.11/116728630/C:p.P1178R:n.C3531G:c.cCt/cGt:SIFTprediction.deleterious:PolyPhenScore.0.629	hits the kinase protein QSK outside its kinase domain
SW480	ENSP00000398470:map.15/40477831/A:p.R363Q:n.G1242A:c.cGa/cAa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein BUBR1 outside its kinase domain
SW480	ENSP00000400312:map.15/75130093/C:p.K445R:n.A1426G:c.aAg/aGg:SIFTprediction.tolerated:PolyPhenScore.0.098	hits the kinase protein ULK3 outside its kinase domain
SW480	ENSP00000408695:map.17/64783081/A:p.V568I:n.G1728A:c.Gtc/Atc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase domain of PKCa
SW480	ENSP00000423665:map.1/205495233/T:p.T196M:n.C807T:c.aCg/aTg:SIFTprediction.deleterious:PolyPhenScore.0.423	hits the kinase domain of PCTAIRE3
SW480	ENSP00000433548:map.12/990912/C:p.T1554P:n.A4660C:c.Acc/Ccc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein Wnk1 outside its kinase domain
SW480	ENSP00000433548:map.12/994487/C,A:p.C2004S:n.G6011C:c.tGc/tCc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein Wnk1 outside its kinase domain
SW480	ENSP00000433548:map.12/994487/C,A:p.C2004Y:n.G6011A:c.tGc/tAc:SIFTprediction.deleterious:PolyPhenScore.0	hits the kinase protein Wnk1 outside its kinase domain
SW480	ENSP00000433548:map.12/998365/T:p.M2306I:n.G6918T:c.atG/atT:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein Wnk1 outside its kinase domain
SW620	ENSP00000166244:map.1/22920150/G:p.Q525R:n.A1646G:c.cAg/cGg:SIFTprediction.tolerated:PolyPhenScore.0.15	hits the kinase protein EphA8 outside its kinase domain
SW620	ENSP00000216911:map.20/54961463/C:p.I57V:n.A437G:c.Att/Gtt:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein AurA outside its kinase domain
SW620	ENSP00000220751:map.8/90784009/G:p.P228A:n.C996G:c.Cct/Gct:SIFTprediction.deleterious:PolyPhenScore.1	hits the kinase domain of RIPK2
SW620	ENSP00000226094:map.17/66533655/G:p.L530S:n.T1877C:c.tTg/tCg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein FAM20A outside its kinase domain
SW620	ENSP00000233027:map.3/52797634/C:p.P225A:n.C876G:c.Cca/Gca:SIFTprediction.deleterious:PolyPhenScore.0.784	hits the kinase domain of NEK4
SW620	ENSP00000240361:map.17/56659018/T:p.G1088D:n.G3349A:c.gGt/gAt:SIFTprediction.tolerated:PolyPhenScore.0.033	hits the kinase protein SgK307 outside its kinase domain
SW620	ENSP00000241453:map.13/28624294/A:p.T227M:n.C762T:c.aCg/aTg:SIFTprediction.deleterious:PolyPhenScore.0.803	hits the kinase protein FLT3 outside its kinase domain
SW620	ENSP00000256443:map.5/68530807/T:p.A2V:n.C108T:c.gCt/gTt:SIFTprediction.deleterious:PolyPhenScore.0.024	hits the kinase protein CDK7 outside its kinase domain
SW620	ENSP00000256458:map.3/10276163/A:p.D431E:n.T1383A:c.gaT/gaA:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase domain of IRAK2
SW620	ENSP00000259750:map.6/43230970/C:p.G623A:n.G1951C:c.gGc/gCc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein TTBK1 outside its kinase domain
SW620	ENSP00000260404:map.15/40564576/T:p.P337L:n.C1436T:c.cCg/cTg:SIFTprediction.deleterious:PolyPhenScore.0.018	hits the kinase protein PAK6 outside its kinase domain
SW620	ENSP00000261170:map.12/14829893/C:p.F281L:n.T980G:c.ttT/ttG:SIFTprediction.tolerated:PolyPhenScore.0.02	hits the kinase protein HSER outside its kinase domain
SW620	ENSP00000261233:map.12/66605228/G:p.I147V:n.A860G:c.Ata/Gta:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein IRAK3 outside its kinase domain
SW620	ENSP00000261937:map.5/180046344/C:p.H890Q:n.C2749G:c.caC/caG:SIFTprediction.deleterious:PolyPhenScore.0.32	hits the kinase domain of FLT4
SW620	ENSP00000262811:map.19/18255359/A:p.G861S:n.G2581A:c.Ggc/Agc:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein MAST3 outside its kinase domain
SW620	ENSP00000262848:map.X/3544520/G:p.I252T:n.T1110C:c.aTt/aCt:SIFTprediction.tolerated:PolyPhenScore.1	hits the kinase domain of PRKX
SW620	ENSP00000263026:map.16/22269867/G:p.Q361R:n.A1556G:c.cAa/cGa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein eEF2K outside its kinase domain
SW620	ENSP00000263791:map.15/40265799/G:p.E556G:n.A1710G:c.gAa/gGa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein GCN2 outside its kinase domain
SW620	ENSP00000263923:map.4/55972974/A:p.Q472H:n.A1712T:c.caA/caT:SIFTprediction.tolerated:PolyPhenScore.0.012	hits the kinase protein KDR outside its kinase domain
SW620	ENSP00000263923:map.4/55979558/T:p.V297I:n.G1185A:c.Gta/Ata:SIFTprediction.tolerated:PolyPhenScore.0.999	hits the kinase protein KDR outside its kinase domain
SW620	ENSP00000263955:map.2/197002262/T:p.I343N:n.T1315A:c.aTc/aAc:SIFTprediction.deleterious:PolyPhenScore.0.728	hits the kinase protein DRAK2 outside its kinase domain
SW620	ENSP00000264316:map.4/48115264/T:p.R45H:n.G220A:c.cGt/cAt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein TXK outside its kinase domain
SW620	ENSP00000265944:map.10/26355906/A:p.R319H:n.G1122A:c.cGt/cAt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MYO3A outside its kinase domain
SW620	ENSP00000265944:map.10/26355992/G:p.I348V:n.A1208G:c.Att/Gtt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MYO3A outside its kinase domain
SW620	ENSP00000265944:map.10/26357748/A:p.V369I:n.G1271A:c.Gtc/Atc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MYO3A outside its kinase domain
SW620	ENSP00000265944:map.10/26446312/A:p.S956N:n.G3033A:c.aGt/aAt:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein MYO3A outside its kinase domain
SW620	ENSP00000265944:map.10/26463043/T:p.T1284S:n.A4016T:c.Act/Tct:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein MYO3A outside its kinase domain
SW620	ENSP00000265944:map.10/26463130/A:p.R1313S:n.C4103A:c.Cgt/Agt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MYO3A outside its kinase domain
SW620	ENSP00000270162:map.21/44837555/A:p.A615V:n.C1977T:c.gCc/gTc:SIFTprediction.deleterious:PolyPhenScore.0.031	hits the kinase protein SIK outside its kinase domain
SW620	ENSP00000270162:map.21/44846016/T:p.G15S:n.G176A:c.Ggt/Agt:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein SIK outside its kinase domain
SW620	ENSP00000275815:map.7/143088867/C:p.M900V:n.A2785G:c.Atg/Gtg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein EphA1 outside its kinase domain
SW620	ENSP00000275815:map.7/143097100/G:p.V160A:n.T566C:c.gTg/gCg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein EphA1 outside its kinase domain
SW620	ENSP00000278616:map.11/108183167/G:p.N1983S:n.A6333G:c.aAt/aGt:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein ATM outside its kinase domain
SW620	ENSP00000283109:map.5/96503523/T:p.G349R:n.G1114A:c.Ggg/Agg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein RIOK2 outside its kinase domain
SW620	ENSP00000283109:map.5/96513471/C:p.S96C:n.C356G:c.tCt/tGt:SIFTprediction.deleterious:PolyPhenScore.0.987	hits the kinase domain of RIOK2
SW620	ENSP00000288135:map.4/55593464/C:p.M541L:n.A1718C:c.Atg/Ctg:SIFTprediction.tolerated:PolyPhenScore.0.008	hits the kinase protein KIT outside its kinase domain
SW620	ENSP00000291281:map.19/47177913/G:p.V835A:n.T3261C:c.gTg/gCg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein PKD2 outside its kinase domain
SW620	ENSP00000291823:map.19/40886993/T:p.R302Q:n.G1190A:c.cGg/cAg:SIFTprediction.tolerated:PolyPhenScore.0.007	hits the kinase domain of HIPK4
SW620	ENSP00000295709:map.2/219537590/T:p.G13V:n.G317T:c.gGc/gTc:SIFTprediction.deleterious:PolyPhenScore.1	hits the kinase domain of Fused
SW620	ENSP00000295709:map.2/219553468/T:p.R477W:n.C1708T:c.Cgg/Tgg:SIFTprediction.deleterious:PolyPhenScore.0.975	hits the kinase protein Fused outside its kinase domain
SW620	ENSP00000295709:map.2/219555262/A:p.R583Q:n.G2027A:c.cGg/cAg:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein Fused outside its kinase domain
SW620	ENSP00000295709:map.2/219562675/A:p.G1003D:n.G3287A:c.gGt/gAt:SIFTprediction.tolerated:PolyPhenScore.0.06	hits the kinase protein Fused outside its kinase domain
SW620	ENSP00000295709:map.2/219563602/A:p.R1112Q:n.G3614A:c.cGg/cAg:SIFTprediction.tolerated:PolyPhenScore.0.004	hits the kinase protein Fused outside its kinase domain
SW620	ENSP00000296084:map.3/133926324/T:p.R210K:n.G629A:c.aGa/aAa:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein RYK outside its kinase domain
SW620	ENSP00000297293:map.7/97822115/A:p.L780M:n.T2631A:c.Ttg/Atg:SIFTprediction.tolerated:PolyPhenScore.0.34	hits the kinase protein LMR2 outside its kinase domain
SW620	ENSP00000298910:map.12/40619082/A:p.R50H:n.G207A:c.cGc/cAc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein LRRK2 outside its kinase domain
SW620	ENSP00000298910:map.12/40707778/A:p.R1514Q:n.G4599A:c.cGa/cAa:SIFTprediction.tolerated:PolyPhenScore.0.043	hits the kinase protein LRRK2 outside its kinase domain
SW620	ENSP00000301178:map.19/41743861/G:p.N266D:n.A986G:c.Aac/Gac:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AXL outside its kinase domain
SW620	ENSP00000301831:map.3/41756965/T:p.V851I:n.G3014A:c.Gta/Ata:SIFTprediction.tolerated:PolyPhenScore.0.01	hits the kinase protein ULK4 outside its kinase domain
SW620	ENSP00000301831:map.3/41756986/T:p.L844M:n.T2993A:c.Ttg/Atg:SIFTprediction.tolerated:PolyPhenScore.0.004	hits the kinase protein ULK4 outside its kinase domain
SW620	ENSP00000301831:map.3/41831203/T:p.A715T:n.G2606A:c.Gct/Act:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein ULK4 outside its kinase domain
SW620	ENSP00000301831:map.3/41841716/C:p.S640A:n.T2381G:c.Tcc/Gcc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein ULK4 outside its kinase domain
SW620	ENSP00000301831:map.3/41877414/C:p.K569R:n.A2169G:c.aAa/aGa:SIFTprediction.tolerated:PolyPhenScore.0.284	hits the kinase protein ULK4 outside its kinase domain
SW620	ENSP00000301831:map.3/41925398/T:p.A542T:n.G2087A:c.Gct/Act:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein ULK4 outside its kinase domain
SW620	ENSP00000301831:map.3/41952852/C:p.S348G:n.A1505G:c.Agt/Ggt:SIFTprediction.deleterious:PolyPhenScore.0.734	hits the kinase protein ULK4 outside its kinase domain
SW620	ENSP00000301831:map.3/41960006/C:p.I224V:n.A1133G:c.Att/Gtt:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase domain of ULK4
SW620	ENSP00000301831:map.3/41996136/C:p.K39R:n.A579G:c.aAa/aGa:SIFTprediction.tolerated:PolyPhenScore.0.111	hits the kinase domain of ULK4
SW620	ENSP00000306678:map.11/113266821/A:p.A239T:n.G809A:c.Gcg/Acg:SIFTprediction.tolerated:PolyPhenScore.0.324	hits the kinase domain of SgK288
SW620	ENSP00000306678:map.11/113270015/C:p.G442R:n.G1418C:c.Ggc/Cgc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein SgK288 outside its kinase domain
SW620	ENSP00000306678:map.11/113270828/A:p.E713K:n.G2231A:c.Gag/Aag:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein SgK288 outside its kinase domain
SW620	ENSP00000307235:map.2/88874891/A:p.A704S:n.G2412T:c.Gct/Tct:SIFTprediction.tolerated:PolyPhenScore.0.005	hits the kinase domain of PEK
SW620	ENSP00000307235:map.2/88895123/C:p.Q166R:n.A799G:c.cAa/cGa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein PEK outside its kinase domain
SW620	ENSP00000308413:map.11/67202156/T:p.A420V:n.C1304T:c.gCc/gTc:SIFTprediction.tolerated:PolyPhenScore.0.01	hits the kinase protein p70S6Kb outside its kinase domain
SW620	ENSP00000309230:map.15/77450964/T:p.R1071K:n.G3491A:c.aGg/aAg:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein SgK269 outside its kinase domain
SW620	ENSP00000310722:map.8/57026229/A:p.A105S:n.G313T:c.Gct/Tct:SIFTprediction.deleterious:PolyPhenScore.0.999	hits the kinase domain of MOS
SW620	ENSP00000313420:map.8/48802984/C:p.A1301G:n.C3959G:c.gCc/gGc:SIFTprediction.deleterious:PolyPhenScore.0.761	hits the kinase protein DNAPK outside its kinase domain
SW620	ENSP00000317985:map.2/11359120/T:p.T431N:n.C1741A:c.aCt/aAt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein ROCK2 outside its kinase domain
SW620	ENSP00000319192:map.7/43664280/G:p.K362E:n.A1263G:c.Aag/Gag:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein DRAK1 outside its kinase domain
SW620	ENSP00000322323:map.7/299881/G:p.N564D:n.A1921G:c.Aac/Gac:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein FAM20C outside its kinase domain
SW620	ENSP00000324560:map.12/132403161/G:p.T816A:n.A2797G:c.Act/Gct:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein ULK1 outside its kinase domain
SW620	ENSP00000332454:map.21/43161357/C:p.M666V:n.A2061G:c.Atg/Gtg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein ANKRD3 outside its kinase domain

SW620	ENSP00000335347:map.14/103934488/C:p.F433S:n.T1298C:c.tTt/tCt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MARK3 outside its kinase domain
SW620	ENSP00000337451:map.3/89521664/A:p.R914H:n.G2966A:c.cGc/cAc:SIFTprediction.tolerated:PolyPhenScore.0.017	hits the kinase protein EphA3 outside its kinase domain
SW620	ENSP00000337451:map.3/89521693/C:p.W924R:n.T2995C:c.Tgg/Cgg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein EphA3 outside its kinase domain
SW620	ENSP00000339299:map.5/14406071/A:p.A1611T:n.G4855A:c.Gct/Act:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein Trio outside its kinase domain
SW620	ENSP00000342105:map.12/68051420/A:p.H245N:n.C1146A:c.Cac/Aac:SIFTprediction.tolerated:PolyPhenScore.0.017	hits the kinase domain of DYRK2
SW620	ENSP00000345083:map.17/21202191/A:p.P40T:n.C367A:c.Ccc/Acc:SIFTprediction.deleterious:PolyPhenScore.0.904	hits the kinase protein MAP2K3 outside its kinase domain
SW620	ENSP00000345083:map.17/21202237/C:p.R55T:n.G413C:c.aGa/aCa:SIFTprediction.tolerated:PolyPhenScore.0.032	hits the kinase protein MAP2K3 outside its kinase domain
SW620	ENSP00000345083:map.17/21203893/C:p.S68P:n.T451C:c.Tca/Cca:SIFTprediction.tolerated:PolyPhenScore.0.065	hits the kinase domain of MAP2K3
SW620	ENSP00000345083:map.17/21203941/A:p.A84T:n.G499A:c.Gcc/Acc:SIFTprediction.tolerated:PolyPhenScore.0.371	hits the kinase domain of MAP2K3
SW620	ENSP00000345083:map.17/21204187/T:p.R94L:n.G530T:c.cGg/cTg:SIFTprediction.deleterious:PolyPhenScore.0.662	hits the kinase domain of MAP2K3
SW620	ENSP00000345083:map.17/21204192/T:p.R96W:n.C535T:c.Cgg/Tgg:SIFTprediction.deleterious:PolyPhenScore.0.996	hits the kinase domain of MAP2K3
SW620	ENSP00000345083:map.17/21207834/T:p.T222M:n.C914T:c.aCg/aTg:SIFTprediction.deleterious:PolyPhenScore.1	hits the kinase domain of MAP2K3
SW620	ENSP00000345083:map.17/21215557/A:p.R293H:n.G1127A:c.cGt/cAt:SIFTprediction.tolerated:PolyPhenScore.0.874	hits the kinase domain of MAP2K3
SW620	ENSP00000345083:map.17/21217513/A:p.V339M:n.G1264A:c.Gtg/Atg:SIFTprediction.deleterious:PolyPhenScore.0.807	hits the kinase protein MAP2K3 outside its kinase domain
SW620	ENSP00000345133:map.11/64597506/C:p.Q1135R:n.A3404G:c.cAg/cGg:SIFTprediction.tolerated:PolyPhenScore.0.008	hits the kinase protein DMPK2 outside its kinase domain
SW620	ENSP00000345629:map.X/19482476/T:p.A192T:n.G574A:c.Gct/Act:SIFTprediction.tolerated:PolyPhenScore.0.408	hits the kinase protein MAP3K7 outside its kinase domain
SW620	ENSP00000346667:map.X/54276067/T:p.S905N:n.G3153A:c.aGt/aAt:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein Wnk3 outside its kinase domain
SW620	ENSP00000348132:map.7/23757162/C:p.Q71H:n.G332C:c.caG/caC:SIFTprediction.tolerated:PolyPhenScore.0.006	hits the kinase protein SgK396 outside its kinase domain
SW620	ENSP00000348132:map.7/23775454/A:p.E261K:n.G900A:c.Gag/Aag:SIFTprediction.tolerated:PolyPhenScore.0.034	hits the kinase protein SgK396 outside its kinase domain
SW620	ENSP00000348132:map.7/23775477/T:p.K268N:n.G923T:c.aaG/aaT:SIFTprediction.deleterious:PolyPhenScore.0.799	hits the kinase protein SgK396 outside its kinase domain
SW620	ENSP00000348132:map.7/23811795/G:p.N621K:n.T1982G:c.aaT/aaG:SIFTprediction.deleterious:PolyPhenScore.0.713	hits the kinase protein SgK396 outside its kinase domain
SW620	ENSP00000348472:map.3/58395863/G:p.K481R:n.A1551G:c.aAg/aGg:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein Slob outside its kinase domain
SW620	ENSP00000348472:map.3/58410554/T:p.A535V:n.C1713T:c.gCa/gTa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein Slob outside its kinase domain
SW620	ENSP00000350195:map.1/27687466/T:p.N622K:n.C2135A:c.aaC/aaA:SIFTprediction.tolerated:PolyPhenScore.0.013	hits the kinase protein MAP3K6 outside its kinase domain
SW620	ENSP00000350195:map.1/27688633/A:p.T455I:n.C1633T:c.aCc/aTc:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein MAP3K6 outside its kinase domain
SW620	ENSP00000353452:map.3/123451773/C:p.L496V:n.C1768G:c.Ctg/Gtg:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein smMLCK outside its kinase domain
SW620	ENSP00000353452:map.3/123457893/A:p.P147S:n.C721T:c.Cca/Tca:SIFTprediction.tolerated:PolyPhenScore.0.011	hits the kinase protein smMLCK outside its kinase domain
SW620	ENSP00000354671:map.1/46476587/G:p.D388E:n.T1447G:c.gaT/gaG:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MAST2 outside its kinase domain
SW620	ENSP00000354671:map.1/46493460/G:p.I659M:n.T2260G:c.aT/atG:SIFTprediction.tolerated:PolyPhenScore.0.491	hits the kinase domain of MAST2
SW620	ENSP00000354877:map.17/19713740/T:p.V370M:n.G1627A:c.Gtg/Atg:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein ULK2 outside its kinase domain
SW620	ENSP00000354991:map.18/56149099/C:p.I2157V:n.A6683G:c.Ata/Gta:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AlphaK2 outside its kinase domain
SW620	ENSP00000354991:map.18/56202768/A:p.A1551S:n.G4865T:c.Gct/Tct:SIFTprediction.deleterious:PolyPhenScore.0.258	hits the kinase protein AlphaK2 outside its kinase domain
SW620	ENSP00000354991:map.18/56203074/A:p.P1449S:n.C4559T:c.Ccg/Tcg:SIFTprediction.deleterious:PolyPhenScore.0.904	hits the kinase protein AlphaK2 outside its kinase domain
SW620	ENSP00000354991:map.18/56204671/T:p.N916K:n.T2962A:c.aaT/aaA:SIFTprediction.deleterious:PolyPhenScore.0.731	hits the kinase protein AlphaK2 outside its kinase domain
SW620	ENSP00000354991:map.18/56204747/A:p.T891I:n.C2886T:c.aCc/aTc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AlphaK2 outside its kinase domain
SW620	ENSP00000354991:map.18/56204932/G:p.K829N:n.A2701C:c.aaA/aaC:SIFTprediction.deleterious:PolyPhenScore.0.824	hits the kinase protein AlphaK2 outside its kinase domain
SW620	ENSP00000354991:map.18/56204945/G:p.R825T:n.G2688C:c.aGa/aCa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AlphaK2 outside its kinase domain
SW620	ENSP00000354991:map.18/56204991/T:p.G810S:n.G2642A:c.Ggt/Agt:SIFTprediction.deleterious:PolyPhenScore.0.951	hits the kinase protein AlphaK2 outside its kinase domain
SW620	ENSP00000354991:map.18/56205262/C:p.H719Q:n.T2371G:c.caT/caG:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AlphaK2 outside its kinase domain
SW620	ENSP00000354991:map.18/56247600/A:p.R136S:n.G622T:c.agG/agT:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein AlphaK2 outside its kinase domain
SW620	ENSP00000355304:map.14/102695693/C:p.Q398R:n.A1425G:c.cAg/cGg:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein MOK outside its kinase domain
SW620	ENSP00000355884:map.1/220808825/T:p.Q410H:n.G1496T:c.caG/caT:SIFTprediction.tolerated:PolyPhenScore.0.94	hits the kinase protein MARK1 outside its kinase domain
SW620	ENSP00000355966:map.1/211840498/C:p.N354S:n.A1200G:c.aAc/aGc:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein NEK2 outside its kinase domain
SW620	ENSP00000356087:map.1/206669465/T:p.P713L:n.C2511T:c.cCt/cTt:SIFTprediction.tolerated:PolyPhenScore.0.008	hits the kinase protein IKKe outside its kinase domain
SW620	ENSP00000356130:map.1/205130413/G:p.C641R:n.T1952C:c.Tgt/Cgt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein Dusty outside its kinase domain
SW620	ENSP00000356530:map.1/182551337/C:p.D541E:n.T1877G:c.gaT/gaG:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase domain of RNAseL
SW620	ENSP00000356530:map.1/182554557/T:p.R462Q:n.G1639A:c.cGa/cAa:SIFTprediction.deleterious:PolyPhenScore.0.85	hits the kinase domain of RNAseL
SW620	ENSP00000356745:map.1/169823718/C:p.Q567R:n.A1915G:c.cAa/cGa:SIFTprediction.tolerated:PolyPhenScore.0.116	hits the kinase protein SCYL3 outside its kinase domain
SW620	ENSP00000357494:map.6/117622184/C:p.S2229C:n.C6885G:c.tCc/tGc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein ROS outside its kinase domain
SW620	ENSP00000357494:map.6/117622188/G:p.K2228Q:n.A6881C:c.Aag/Cag:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein ROS outside its kinase domain
SW620	ENSP00000357494:map.6/117622233/T:p.D2213N:n.G6836A:c.Gac/Aac:SIFTprediction.tolerated:PolyPhenScore.0.009	hits the kinase domain of ROS
SW620	ENSP00000357615:map.6/116325142/T:p.G122R:n.G811A:c.Gga/Aga:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein FRK outside its kinase domain
SW620	ENSP00000359424:map.10/101977883/T:p.V268I:n.G857A:c.Gta/Ata:SIFTprediction.tolerated:PolyPhenScore.0.005	hits the kinase domain of IKKa
SW620	ENSP00000361025:map.9/136270538/G:p.L679R:n.T2143G:c.cTg/cCg:SIFTprediction.deleterious:PolyPhenScore.0.912	hits the kinase protein SgK071 outside its kinase domain
SW620	ENSP00000362139:map.1/38185723/T:p.R807Q:n.G2420A:c.cGg/cAg:SIFTprediction.deleterious:PolyPhenScore.0.792	hits the kinase domain of EphA10
SW620	ENSP00000362139:map.1/38188740/T:p.V645I:n.G1933A:c.Gtc/Atc:SIFTprediction.tolerated:PolyPhenScore.0.005	hits the kinase domain of EphA10
SW620	ENSP00000362139:map.1/38188787/G:p.L629P:n.T1886C:c.cTg/cCg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein EphA10 outside its kinase domain
SW620	ENSP00000362139:map.1/38227086/T:p.F281I:n.T841A:c.Ttc/Atc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein EphA10 outside its kinase domain
SW620	ENSP00000363708:map.2/203420712/A:p.S775N:n.G2863A:c.aGc/aAc:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein BMPR2 outside its kinase domain
SW620	ENSP00000364204:map.1/20977000/C:p.N521T:n.A1656C:c.aAt/aCt:SIFTprediction.deleterious:PolyPhenScore.0.007	hits the kinase protein PINK1 outside its kinase domain
SW620	ENSP00000364361:map.2/174128513/T:p.S531L:n.C1670T:c.tCg/tTg:SIFTprediction.tolerated:PolyPhenScore.0.001	hits the kinase protein ZAK outside its kinase domain
SW620	ENSP00000364860:map.9/94495608/C:p.T245A:n.A932G:c.Aca/Gca:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein ROR2 outside its kinase domain
SW620	ENSP00000366488:map.9/71628207/C:p.H268D:n.C833G:c.Cat/Gat:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase domain of PKACg
SW620	ENSP00000369375:map.9/27183463/C:p.Q346P:n.A1479C:c.cAg/cCg:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein TIE2 outside its kinase domain
SW620	ENSP00000372035:map.13/21562832/T:p.G363S:n.G1493A:c.Ggc/Agc:SIFTprediction.tolerated:PolyPhenScore.0.005	hits the kinase protein LATS2 outside its kinase domain
SW620	ENSP00000372035:map.13/21562948/A:p.A324V:n.C1377T:c.gCg/gTg:SIFTprediction.tolerated:PolyPhenScore.0.031	hits the kinase protein LATS2 outside its kinase domain
SW620	ENSP00000373600:map.15/101606889/A:p.G1938D:n.G6172A:c.gGc/gAc:SIFTprediction.tolerated:PolyPhenScore.0.013	hits the kinase protein LRRK1 outside its kinase domain
SW620	ENSP00000375986:map.6/161469774/A:p.R157H:n.G618A:c.cGt/cAt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MAP3K4 outside its kinase domain
SW620	ENSP00000378288:map.16/46773999/A:p.V180L:n.G654T:c.Gtg/Ttg:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein caMLCK outside its kinase domain
SW620	ENSP00000378945:map.4/82065465/T:p.G392S:n.G1291A:c.Ggt/Agt:SIFTprediction.tolerated:PolyPhenScore.0.999	hits the kinase protein PKG2 outside its kinase domain
SW620	ENSP00000380066:map.19/39100236/A:p.R336W:n.C1114T:c.Cgg/Tgg:SIFTprediction.deleterious:PolyPhenScore.0.391	hits the kinase protein HPK1 outside its kinase domain
SW620	ENSP00000381129:map.4/2990499/T:p.R65L:n.G537T:c.cGt/cTt:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein GPRK4 outside its kinase domain
SW620	ENSP00000381129:map.4/3006043/T:p.A142V:n.C768T:c.gCc/gTc:SIFTprediction.tolerated:PolyPhenScore.0.004	hits the kinase protein GPRK4 outside its kinase domain
SW620	ENSP00000381129:map.4/3015553/A:p.V247I:n.G1082A:c.Gta/Ata:SIFTprediction.tolerated:PolyPhenScore.0.669	hits the kinase domain of GPRK4
SW620	ENSP00000382423:map.5/56177443/A:p.D806N:n.G2416A:c.Gat/Aat:SIFTprediction.tolerated:PolyPhenScore.0.111	hits the kinase protein MAP3K1 outside its kinase domain
SW620	ENSP00000382423:map.5/56177743/A:p.V906I:n.G2716A:c.Gtc/Atc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein MAP3K1 outside its kinase domain
SW620	ENSP00000382544:map.22/19119751/T:p.T280M:n.C1431T:c.aCg/aTg:SIFTprediction.tolerated:PolyPhenScore.0.003	hits the kinase protein TSSK2 outside its kinase domain
SW620	ENSP00000383234:map.18/48190440/A:p.V38M:n.G1112A:c.Gtg/Atg:SIFTprediction.tolerated:PolyPhenScore.0.357	hits the kinase domain of ERK4
SW620	ENSP00000384442:map.1/1650787/C:p.H112R:n.A415G:c.cAt/cGt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein CDK11b outside its kinase domain
SW620	ENSP00000384442:map.1/1650797/G:p.C109R:n.T405C:c.Tgt/Cgt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein CDK11b outside its kinase domain
SW620	ENSP00000384442:map.1/1650832/G:p.V97A:n.T370C:c.gTt/gCt:SIFTprediction.tolerated:PolyPhenScore.0.013	hits the kinase protein CDK11b outside its kinase domain
SW620	ENSP00000384442:map.1/1650845/A:p.R93W:n.C357T:c.Cgg/Tgg:SIFTprediction.deleterious:PolyPhenScore.0.999	hits the kinase protein CDK11b outside its kinase domain
SW620	ENSP00000386213:map.2/171260787/A:p.V770I:n.G2451A:c.Gta/Ata:SIFTprediction.tolerated:PolyPhenScore.0.004	hits the kinase protein MYO3B outside its kinase domain
SW620	ENSP00000386213:map.2/171356274/A:p.R1082K:n.G3388A:c.aGg/aAg:SIFTprediction.tolerated:PolyPhenScore.0.005	hits the kinase protein MYO3B outside its kinase domain
SW620	ENSP00000386456:map.2/69741854/G:p.K509Q:n.A1902C:c.Aaa/Caa:SIFTprediction.tolerated:PolyPhenScore.0.002	hits the kinase protein AAK1 outside its kinase domain
SW620	ENSP00000389015:map.19/56041255/G:p.A298P:n.G930C:c.Gcc/Ccc:SIFTprediction.tolerated:PolyPhenScore.0.004	hits the kinase domain of SgK069
SW620	ENSP00000389015:map.19/56047448/G:p.C72R:n.T252C:c.Tgc/Cgc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase domain of SgK069
SW620	ENSP00000391295:map.11/116728630/C:p.P1178R:n.C3531G:c.cCt/cGt:SIFTprediction.deleterious:PolyPhenScore.0.629	hits the kinase protein QSK outside its kinase domain
SW620	ENSP00000398470:map.15/40477831/A:p.R363Q:n.G1242A:c.cGa/cAa:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein BUBR1 outside its kinase domain
SW620	ENSP00000400312:map.15/75130093/C:p.K445R:n.A1426G:c.aAg/aGg:SIFTprediction.tolerated:PolyPhenScore.0.098	hits the kinase protein ULK3 outside its kinase domain
SW620	ENSP00000408695:map.17/64783081/A:p.V568I:n.G1728A:c.Gtc/Atc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase domain of PKCa
SW620	ENSP00000423665:map.1/205495233/T:p.T196M:n.C807T:c.aCg/aTg:SIFTprediction.deleterious:PolyPhenScore.0.423	hits the kinase domain of PCTAIRE3
SW620	ENSP00000427235:map.4/151177340/T:p.P747S:n.C2239T:c.Cct/Tct:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein DCLK2 outside its kinase domain
SW620	ENSP00000427235:map.4/151177341/G:p.P747R:n.C2240G:c.cCt/cGt:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein DCLK2 outside its kinase domain
SW620	ENSP00000433548:map.12/990912/C:p.T1554P:n.A4660C:c.Acc/Ccc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein Wnk1 outside its kinase domain
SW620	ENSP00000433548:map.12/994487/C:p.C2004S:n.G6011C:c.tGc/tCc:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein Wnk1 outside its kinase domain
SW620	ENSP00000433548:map.12/998365/T:p.M2306I:n.G6918T:c.atG/atT:SIFTprediction.tolerated:PolyPhenScore.0	hits the kinase protein Wnk1 outside its kinase domain

Supplementary Table 2

RefSeq Accession Number	Gene Symbol	GeneID
NM_000020	ACVRL1	94
NM_000051	ATM	472
NM_000061	BTK	695
NM_000075	CDK4	1019
NM_000142	FGFR3	2261
NM_000180	GUCY2D	3000
NM_000208	INSR	3643
NM_000215	JAK3	3718
NM_000222	KIT	3815
NM_000245	MET	4233
NM_000294	PHKG2	5261
NM_000314	PTEN	5728
NM_000321	RB1	5925
NM_000455	STK11	6794
NM_000459	TEK	7010
NM_000546	TP53	7157
NM_000548	TSC2	7249
NM_000875	IGF1R	3480
NM_000906	NPR1	4881
NM_000907	NPR2	4882
NM_000932	PLCB3	5331
NM_000933	PLCB4	5332
NM_001001671	FLJ16518	389840
NM_001001716	NFKBIB	4793
NM_001001852	PIM3	415116
NM_001003786	LYK5	92335
NM_001003787	LYK5	92335
NM_001003788	LYK5	92335
NM_001004023	DYRK3	8444
NM_001004105	GRK6	2870
NM_001005862	ERBB2	2064
NM_001005915	ERBB3	2065
NM_001006665	RPS6KA1	6195
NM_001006932	RPS6KA2	6196
NM_001006943	EPHA8	2046
NM_001006944	RPS6KA4	8986
NM_001007071	RPS6KB2	6199
NM_001007156	NTRK3	4916
NM_001007792	NTRK1	4914
NM_001008910	STK16	8576
NM_001009565	CDKL4	344387
NM_001011664	CSNK1G1	53944
NM_001012331	NTRK1	4914
NM_001012418	LOC340156	340156
NM_001013703	EIF2AK4	440275
NM_001014431	AKT1	207
NM_001014796	DDR2	4921
NM_001014833	PAK4	10298
NM_001015878	AURKC	6795
NM_001018046	FLJ23074	80122
NM_001018066	NTRK2	4915
NM_001024401	SBK1	388228
NM_001024660	HAPIP	8997
NM_001024847	TGFBR2	7048
NM_001025105	CSNK1A1	1452
NM_001025242	IRAK1	3654
NM_001025243	IRAK1	3654
NM_001025778	VRK3	51231
NM_001031741	FLJ32685	152110
NM_001031812	CSNK1G3	1456
NM_001032296	STK24	8428
NM_001033582	PRKCZ	5590
NM_001037343	CDKL5	6792
NM_001079	ZAP70	7535
NM_001080395	custom	
NM_001105	ACVR1	90
NM_001106	ACVR2B	93
NM_001184	ATR	545
NM_001203	BMPR1B	658
NM_001204	BMPR2	659
NM_001211	BUB1B	701
NM_001237	CCNA2	890
NM_001238	CCNE1	898
NM_001258	CDK3	1018
NM_001259	CDK6	1021
NM_001260	CDK8	1024
NM_001261	CDK9	1025
NM_001274	CHEK1	1111
NM_001278	CHUK	1147
NM_001292	CLK3	1198
NM_001315	MAPK14	1432
NM_001319	CSNK1G2	1455
NM_001348	DAPK3	1613
NM_001429	EP300	2033
NM_001433	ERN1	2081
NM_001522	GUCY2F	2986
NM_001556	IKBKB	3551
NM_001570	IRAK2	3656
NM_001616	ACVR2	92
NM_001619	ADRBK1	156
NM_001626	AKT2	208
NM_001654	ARAF1	369
NM_001664	RHOA	387
NM_001699	AXL	558
NM_001715	BLK	640
NM_001726	BRDT	676
NM_001744	CAMK4	814
NM_001759	CCND2	894
NM_001760	CCND3	896
NM_001786	CDC2	983
NM_001791	CDC42	998
NM_001798	CDK2	1017
NM_001799	CDK7	1022
NM_001892	CSNK1A1	1452
NM_001893	CSNK1D	1453
NM_001894	CSNK1E	1454
NM_001896	CSNK2A2	1459
NM_001904	CTNNB1	1499

NM_001949	E2F3	1871
NM_001950	E2F4	1874
NM_001982	ERBB3	2065
NM_002005	FES	2242
NM_002019	FLT1	2321
NM_002020	FLT4	2324
NM_002031	FRK	2444
NM_002082	GRK6	2870
NM_002093	GSK3B	2932
NM_002110	HCK	3055
NM_002227	JAK1	3716
NM_002228	JUN	3725
NM_002253	KDR	3791
NM_002314	LIMK1	3984
NM_002344	LTK	4058
NM_002350	LYN	4067
NM_002376	MARK3	4140
NM_002401	MAP3K3	4215
NM_002419	MAP3K11	4296
NM_002446	MAP3K10	4294
NM_002447	MST1R	4486
NM_002497	NEK2	4751
NM_002503	NFKBIB	4793
NM_002524	NRAS	4893
NM_002530	NTRK3	4916
NM_002576	PAK1	5058
NM_002577	PAK2	5062
NM_002578	PAK3	5063
NM_002595	PCTK2	5128
NM_002596	PCTK3	5129
NM_002609	PDGFRB	5159
NM_002610	PDK1	5163
NM_002611	PDK2	5164
NM_002612	PDK4	5166
NM_002613	PDPK1	5170
NM_002645	PIK3C2A	5286
NM_002646	PIK3C2B	5287
NM_002648	PIM1	5292
NM_002649	PIK3CG	5294
NM_002660	PLCG1	5335
NM_002661	PLCG2	5336
NM_002730	PRKACA	5566
NM_002731	PRKACB	5567
NM_002732	PRKACG	5568
NM_002737	PRKCA	5578
NM_002738	PRKCB1	5579
NM_002739	PRKCG	5582
NM_002740	PRKCI	5584
NM_002741	PKN1	5585
NM_002742	PRKCM	5587
NM_002744	PRKCZ	5590
NM_002745	MAPK1	5594
NM_002746	MAPK3	5595
NM_002747	MAPK4	5596
NM_002748	MAPK6	5597
NM_002750	MAPK8	5599
NM_002751	MAPK11	5600
NM_002752	MAPK9	5601
NM_002754	MAPK13	5603
NM_002755	MAP2K1	5604
NM_002757	MAP2K5	5607
NM_002758	MAP2K6	5608
NM_002759	PRKR	5610
NM_002760	PRKY	5616
NM_002821	PTK7	5754
NM_002880	RAF1	5894
NM_002881	RALB	5899
NM_002929	GRK1	6011
NM_002944	ROS1	6098
NM_002953	RPS6KA1	6195
NM_002958	RYK	6259
NM_002969	MAPK12	6300
NM_003010	MAP2K4	6416
NM_003137	SRPK1	6732
NM_003151	STAT4	6775
NM_003152	STAT5A	6776
NM_003153	STAT6	6778
NM_003157	NEK4	6787
NM_003159	CDKL5	6792
NM_003160	AURKC	6795
NM_003161	RPS6KB1	6198
NM_003177	SYK	6850
NM_003215	TEC	7006
NM_003242	TGFBR2	7048
NM_003318	TTK	7272
NM_003328	TXK	7294
NM_003331	TYK2	7297
NM_003384	VRK1	7443
NM_003390	WEE1	7465
NM_003496	TRRAP	8295
NM_003503	CDC7	8317
NM_003565	ULK1	8408
NM_003576	STK24	8428
NM_003582	DYRK3	8444
NM_003583	DYRK2	8445
NM_003600	STK6	6790
NM_003618	MAP4K3	8491
NM_003656	CAMK1	8536
NM_003674	CDK10	8558
NM_003684	MKNK1	8569
NM_003688	CASK	8573
NM_003691	STK16	8576
NM_003804	RIPK1	8737
NM_003821	RIPK2	8767
NM_003831	RIOK3	8780
NM_003845	DYRK4	8798
NM_003852	TIF1	8805
NM_003913	PRPF4B	8899
NM_003942	RPS6KA4	8986

NM_003948	CDKL2	8999
NM_003952	RPS6KB2	6199
NM_003954	MAP3K14	9020
NM_003957	STK29	9024
NM_003985	TNK1	8711
NM_003992	CLK3	1198
NM_003993	CLK2	1196
NM_003995	NPR2	4882
NM_004071	CLK1	1195
NM_004073	PLK3	1263
NM_004091	E2F2	1870
NM_004119	FLT3	2322
NM_004196	CDKL1	8814
NM_004197	STK19	8859
NM_004203	PKMYT1	9088
NM_004217	AURKB	9212
NM_004226	STK17B	9262
NM_004302	ACVR1B	91
NM_004304	ALK	238
NM_004327	BCR	613
NM_004329	BMPR1A	657
NM_004333	BRAF	673
NM_004336	BUB1	699
NM_004380	CREBBP	1387
NM_004383	CSK	1445
NM_004384	CSNK1G3	1456
NM_004409	DMPK	1760
NM_004422	DVL2	1856
NM_004423	DVL3	1857
NM_004431	EPHA2	1969
NM_004438	EPHA4	2043
NM_004440	EPHA7	2045
NM_004441	EPHB1	2047
NM_004442	EPHB2	2048
NM_004443	EPHB3	2049
NM_004444	EPHB4	2050
NM_004445	EPHB6	2051
NM_004448	ERBB2	2064
NM_004517	ILK	3611
NM_004556	NFKBIE	4794
NM_004560	ROR2	4920
NM_004570	PIK3C2G	5288
NM_004573	PLCB2	5330
NM_004579	MAP4K2	5871
NM_004586	RPS6KA3	6197
NM_004606	TAF1	6872
NM_004612	TGFBR1	7046
NM_004635	MAPKAPK3	7867
NM_004672	MAP3K6	9064
NM_004690	LATS1	9113
NM_004714	DYRK1B	9149
NM_004721	MAP3K13	9175
NM_004734	DCAMKL1	9201
NM_004755	RPS6KA5	9252
NM_004759	MAPKAPK2	9261
NM_004760	STK17A	9263
NM_004836	EIF2AK3	9451
NM_004850	ROCK2	9475
NM_004935	CDK5	1020
NM_004938	DAPK1	1612
NM_004954	MARK2	2011
NM_004958	FRAP1	2475
NM_004963	GUCY2C	2984
NM_004972	JAK2	3717
NM_004985	KRAS2	3845
NM_005012	ROR1	4919
NM_005030	PLK1	5347
NM_005043	MAP2K7	5609
NM_005044	PRKX	5613
NM_005104	BRD2	6046
NM_005109	OSR1	9943
NM_005157	ABL1	25
NM_005158	ABL2	27
NM_005160	ADRBK2	157
NM_005163	AKT1	207
NM_005204	MAP3K8	1326
NM_005211	CSF1R	1436
NM_005225	E2F1	1869
NM_005232	EPHA1	2041
NM_005235	ERBB4	2066
NM_005246	FER	2241
NM_005248	FGR	2268
NM_005252	FOS	2353
NM_005255	GAK	2580
NM_005307	GRK4	2868
NM_005308	GRK5	2869
NM_005356	LCK	3932
NM_005359	SMAD4	4089
NM_005372	MOS	4342
NM_005391	PDK3	5165
NM_005400	PRKCE	5581
NM_005402	RALA	5898
NM_005406	ROCK1	6093
NM_005419	STAT2	6773
NM_005424	TIE	7075
NM_005433	YES1	7525
NM_005465	AKT3	10000
NM_005546	ITK	3702
NM_005569	LIMK2	3985
NM_005592	MUSK	4593
NM_005627	SGK	6446
NM_005633	SOS1	6654
NM_005734	HIPK3	10114
NM_005762	TRIM28	10155
NM_005781	ACK1	10188
NM_005813	PRKCN	23683
NM_005881	BCKDK	10295
NM_005884	PAK4	10298
NM_005906	MAK	4117

NM_005923	MAP3K5	4217
NM_005975	PTK6	5753
NM_005990	STK10	6793
NM_006035	CDC42BPB	9578
NM_006180	NTRK2	4915
NM_006182	DDR2	4921
NM_006206	PDGFRA	5156
NM_006213	PHKG1	5260
NM_006251	PRKAA1	5562
NM_006252	PRKAA2	5563
NM_006255	PRKCH	5583
NM_006256	PKN2	5586
NM_006257	PRKCQ	5588
NM_006258	PRKG1	5592
NM_006259	PRKG2	5593
NM_006281	STK3	6788
NM_006282	STK4	6789
NM_006285	TESK1	7016
NM_006293	TYRO3	7301
NM_006296	VRK2	7444
NM_006301	MAP3K12	7786
NM_006343	MERTK	10461
NM_006374	STK25	10494
NM_006482	DYRK2	8445
NM_006483	DYRK1B	9149
NM_006484	DYRK1B	9149
NM_006575	MAP4K5	11183
NM_006609	MAP3K2	10746
NM_006622	PLK2	10769
NM_006648	PRKWNK2	65268
NM_006712	FASTK	10922
NM_006724	MAP3K4	4216
NM_006742	PSKH1	5681
NM_006852	TLK2	11011
NM_006871	RIPK3	11035
NM_006875	PIM2	11040
NM_006879	MDM2	4193
NM_006880	MDM2	4193
NM_006882	MDM2	4193
NM_006904	PRKDC	5591
NM_006908	RAC1	5879
NM_007064	HAPIP	8997
NM_007118	TRIO	7204
NM_007170	TESK2	10420
NM_007174	CIT	11113
NM_007181	MAP4K1	11184
NM_007194	CHEK2	11200
NM_007199	IRAK3	11213
NM_007271	STK38	11329
NM_007296	BRCA1	672
NM_007301	BRCA1	672
NM_007303	BRCA1	672
NM_007313	ABL1	25
NM_007314	ABL2	27
NM_007315	STAT1	6772
NM_007371	BRD3	8019
NM_012119	CCRK	23552
NM_012224	NEK1	4750
NM_012290	TLK1	9874
NM_012395	PFTK1	5218
NM_012424	RPS6KC1	26750
NM_012448	STAT5B	6777
NM_013233	STK39	27347
NM_013254	TBK1	29110
NM_013302	EEF2K	29904
NM_013355	PKN3	29941
NM_013392	NRBP	29959
NM_013993	DDR1	780
NM_013994	DDR1	780
NM_014002	IKBKE	9641
NM_014006	SMG1	23049
NM_014215	INSRR	3645
NM_014226	RAGE	5891
NM_014238	XM_290793	8844
NM_014264	PLK4	10733
NM_014299	BRD4	23476
NM_014326	DAPK2	23604
NM_014365	HSPB8	26353
NM_014370	STK23	26576
NM_014397	NEK6	10783
NM_014413	HRI	27102
NM_014496	RPS6KA6	27330
NM_014572	LATS2	26524
NM_014586	HUNK	30811
NM_014602	PIK3R4	30849
NM_014683	ULK2	9706
NM_014720	SLK	9748
NM_014791	MELK	9833
NM_014826	CDC42BPA	8476
NM_014840	ARK5	9891
NM_014911	AAK1	22848
NM_014916	LMTK2	22853
NM_014975	MAST1	22983
NM_015000	STK38L	23012
NM_015028	KIAA0551	23043
NM_015076	CDK11	23097
NM_015092	SMG1	23049
NM_015112	MAST2	23139
NM_015148	PASK	23178
NM_015191	SIK2	23235
NM_015375	RIPK5	25778
NM_015518	DKFZP434C131	25989
NM_015690	STK36	27148
NM_015716	MINK	50488
NM_015905	TIF1	8805
NM_015906	TRIM33	51592
NM_015978	TNNI3K	51086
NM_015981	CAMK2A	815
NM_016123	IRAK4	51135

NM_016151	TAO1	9344
NM_016231	NLK	51701
NM_016269	LEF1	51176
NM_016276	SGK2	10110
NM_016281	JIK	51347
NM_016440	VRK3	51231
NM_016457	PRKD2	25865
NM_016507	CRK7	51755
NM_016508	CDKL3	51265
NM_016542	MST4	51765
NM_016653	ZAK	51776
NM_016733	LIMK2	3985
NM_016735	LIMK1	3984
NM_017433	MYO3A	53904
NM_017449	EPHB2	2048
NM_017490	MARK2	2011
NM_017525	HSMDPKIN	55561
NM_017572	MKNK2	2872
NM_017593	BMP2K	55589
NM_017662	TRPM6	140803
NM_017672	TRPM7	54822
NM_017719	SNRK	54861
NM_017771	PXK	54899
NM_017886	FLJ20574	54986
NM_017988	FLJ10074	55681
NM_018343	RIOK2	55781
NM_018401	STK32B	55351
NM_018423	STYK1	55359
NM_018492	TOPK	55872
NM_018571	ALS2CR2	55437
NM_018650	MARK1	4139
NM_018890	RAC1	5879
NM_018979	PRKWNK1	65125
NM_019884	GSK3A	2931
NM_020168	PAK6	56924
NM_020247	CABC1	56997
NM_020328	ACVR1B	91
NM_020341	PAK7	57144
NM_020397	CAMK1D	57118
NM_020421	ADCK1	57143
NM_020439	CAMK1G	57172
NM_020526	EPHA8	2046
NM_020529	NFKBIA	4792
NM_020547	AMHR2	269
NM_020630	RET	5979
NM_020639	RIPK4	54101
NM_020666	CLK4	57396
NM_020680	SCYL1	57410
NM_020761	raptor	57521
NM_020778	MIDORI	57538
NM_020791	KIAA1361	57551
NM_020922	PRKWNK3	65267
NM_021055	TSC2	7249
NM_021056	TSC2	7249
NM_021133	RNASEL	6041
NM_021135	RPS6KA2	6196
NM_021158	TRIB3	57761
NM_021574	BCR	613
NM_021643	TRIB2	28951
NM_021872	CDC25B	994
NM_021913	AXL	558
NM_022051	EGLN1	54583
NM_022740	HIPK2	28996
NM_022963	FGFR4	2264
NM_022965	FGFR3	2261
NM_022972	FGFR2	2263
NM_022975	FGFR2	2263
NM_023031	FGFR2	2263
NM_023106	FGFR1	2260
NM_023110	FGFR1	2260
NM_023111	FGFR1	2260
NM_024046	MGC8407	79012
NM_024652	LRRK1	79705
NM_024876	ADCK4	79934
NM_025052	FLJ23074	80122
NM_025144	LAK	80216
NM_025164	KIAA0999	23387
NM_025195	TRIB1	10221
NM_030662	MAP2K2	5605
NM_030906	STK33	65975
NM_030952	SNARK	81788
NM_031267	CDC2L5	8621
NM_031268	PDPK1	5170
NM_031272	TEX14	56155
NM_031414	STK31	56164
NM_031417	MARK4	57787
NM_031464	RPS6KL1	83694
NM_031965	GSG2	83903
NM_031966	CCNB1	891
NM_031988	MAP2K6	5608
NM_032017	MGC4796	83931
NM_032028	STK22D	83942
NM_032037	SSTK	83983
NM_032237	FLJ23356	84197
NM_032294	CAMKK1	84254
NM_032387	PRKWNK4	65266
NM_032409	PINK1	65018
NM_032430	KIAA1811	84446
NM_032435	KIAA1804	84451
NM_032454	STK19	8859
NM_032538	TTBK1	84630
NM_032844	MASTL	84930
NM_032960	MAPKAPK2	9261
NM_033015	FASTK	10922
NM_033018	PCTK1	5127
NM_033019	PCTK1	5127
NM_033020	TRIM33	51592
NM_033115	MGC16169	93627
NM_033116	NEK9	91754

NM_033118	MYLK2	85366
NM_033126	PSKH2	85481
NM_033141	MAP3K9	4293
NM_033266	ERN2	10595
NM_033360	KRAS2	3845
NM_033379	CDC2	983
NM_033532	CDC2L2	985
NM_033534	CDC2L2	985
NM_033537	CDC2L2	985
NM_033550	TP53RK	112858
NM_044472	CDC42	998
NM_052827	CDK2	1017
NM_052841	STK22C	81629
NM_052843	OBSCN	84033
NM_052853	ADCK2	90956
NM_052947	HAK	115701
NM_052984	CDK4	1019
NM_052987	CDK10	8558
NM_053006	STK22B	23617
NM_053029	MYLK	4638
NM_053030	MYLK	4638
NM_053031	MYLK	4638
NM_053056	CCND1	595
NM_057735	CCNE2	9134
NM_057749	CCNE2	9134
NM_058195	CDKN2A	1029
NM_058197	CDKN2A	1029
NM_058243	BRD4	23476
NM_080823	SRMS	6725
NM_080836	STK35	140901
NM_130436	DYRK1A	1859
NM_130438	DYRK1A	1859
NM_133378	TTN	7273
NM_133379	TTN	7273
NM_133432	TTN	7273
NM_133494	NEK7	140609
NM_133646	ZAK	51776
NM_138293	ATM	472
NM_138370	LOC91461	91461
NM_138923	TAF1	6872
NM_138957	MAPK1	5594
NM_138980	MAPK10	5602
NM_138981	MAPK10	5602
NM_138982	MAPK10	5602
NM_138993	MAPK11	5600
NM_138995	MYO3B	140469
NM_139013	MAPK14	1432
NM_139014	MAPK14	1432
NM_139021	ERK8	225689
NM_139032	MAPK7	5598
NM_139034	MAPK7	5598
NM_139046	MAPK8	5599
NM_139047	MAPK8	5599
NM_139062	CSNK1D	1453
NM_139069	MAPK9	5601
NM_139070	MAPK9	5601
NM_139078	MAPKAPK5	8550
NM_139158	ALS2CR7	65061
NM_139209	GRK7	131890
NM_139266	STAT1	6772
NM_139276	STAT3	6774
NM_139354	MATK	4145
NM_139355	MATK	4145
NM_144610	FLJ25006	124923
NM_144617	HSPB6	126393
NM_144624	KIS	127933
NM_144685	HIPK4	147746
NM_145001	STK32A	202374
NM_145109	MAP2K3	5606
NM_145110	MAP2K3	5606
NM_145161	MAP2K5	5607
NM_145185	MAP2K7	5609
NM_145203	CSNK1A1L	122011
NM_145259	ACVR1C	130399
NM_145319	MAP3K6	9064
NM_145331	MAP3K7	6885
NM_145332	MAP3K7	6885
NM_145333	MAP3K7	6885
NM_145686	MAP4K4	9448
NM_145687	MAP4K4	9448
NM_145862	CHEK2	11200
NM_145906	RIOK3	8780
NM_145910	NEK11	79858
NM_152221	CSNK1E	1454
NM_152461	ERN1	2081
NM_152534	FLJ32685	152110
NM_152619	MGC45428	166614
NM_152649	FLJ34389	197259
NM_152720	NEK3	4752
NM_152756	MGC39830	253260
NM_152835	LOC149420	149420
NM_152881	PTK7	5754
NM_152883	PTK7	5754
NM_153005	RIOK1	83732
NM_153047	FYN	2534
NM_153048	FYN	2534
NM_153361	MGC42105	167359
NM_153498	CAMK1D	57118
NM_153500	CAMKK2	10645
NM_153710	C9orf96	169436
NM_153809	TAF1L	138474
NM_153827	MINK	50488
NM_153831	PTK2	5747
NM_170663	MINK	50488
NM_170693	SGK2	10110
NM_170709	SGKL	23678
NM_171825	CAMK2A	815
NM_172079	CAMK2B	816
NM_172081	CAMK2B	816

NM_172083	CAMK2B	816
NM_172115	CAMK2D	817
NM_172127	CAMK2D	817
NM_172128	CAMK2D	817
NM_172171	CAMK2G	818
NM_172206	CAMKK1	84254
NM_172226	CAMKK2	10645
NM_173174	PTK2B	2185
NM_173176	PTK2B	2185
NM_173354	SNF1LK	150094
NM_173500	TTBK2	146057
NM_173575	STK32C	282974
NM_173598	KSR2	283455
NM_173641	FLJ33655	284656
NM_173655	DKFZp434C1418	285220
NM_173677	FLJ40852	285962
NM_174922	ADCK5	203054
NM_174944	C14orf20	283629
NM_175866	KIS	127933
NM_176795	HRAS	3265
NM_176800	PRPF4B	8899
NM_177559	CSNK2A1	1457
NM_177560	CSNK2A1	1457
NM_177990	PAK7	57144
NM_178170	NEK8	284086
NM_178432	CCRK	23552
NM_178510	ANKK1	255239
NM_178564	LOC340371	340371
NM_181093	PACE-1	57147
NM_181358	HIPK1	204851
NM_181690	AKT3	10000
NM_181870	DVL1	1855
NM_182398	RPS6KA5	9252
NM_182472	EPHA5	2044
NM_182493	LOC91807	91807
NM_182644	EPHA3	2042
NM_182687	PKMYT1	9088
NM_182691	SRPK2	6733
NM_182692	SRPK2	6733
NM_182734	PLCB1	23236
NM_182779	DVL1	1855
NM_182797	PLCB4	5332
NM_182811	PLCG1	5335
NM_182925	FLT4	2324
NM_182982	GRK4	2868
NM_198268	HIPK1	204851
NM_198269	HIPK1	204851
NM_198291	SRC	6714
NM_198393	TEX14	56155
NM_198435	STK6	6790
NM_198437	STK6	6790
NM_198452	PNCK	139728
NM_198465	NRK	203447
NM_198578	LRRK2	120892
NM_198794	MAP4K5	11183
NM_198828	LOC375449	375449
NM_198892	BMP2K	55589
NM_198973	MKNK1	8569
NM_199054	MKNK2	2872
NM_199289	NEK5	341676
NM_199462	RIPK5	25778
NM_201282	EGFR	1956
NM_201284	EGFR	1956
NM_201567	CDC25A	993
NM_203281	BMX	660
NM_203351	MAP3K3	4215
NM_206961	LTK	4058
NM_207189	BRDT	676
NM_207519	ZAP70	7535
NM_207578	PRKACB	5567
NM_212503	PCTK3	5129
NM_212530	CDC25B	994
NM_212535	PRKCB1	5579
NM_212539	PRKCD	5580
NM_213560	PKN1	5585
NM_213662	STAT3	6774
XM_001131586	custom	
XM_001131886	custom	
XM_038150	MAST3	23031
XM_039796	KIAA0551	23043
XM_042066	MAP3K1	4214
XM_047355	KIAA1765	85443
XM_055866	LMTK3	114783
XM_058513	LRRK2	120892
XM_291277	DKFZp761P0423	157285
XM_370878	KIAA2002	79834
XM_380173	MGC39830	253260
XM_496653	DKFZp434C1418	285220

Chapter IV

Epilogue

Concluding Remarks

Chapter IV - Concluding remarks

In this thesis, we have shown some early results of targeting cancer biology using an integrative Network Biology approach. By analyzing different cellular aspects using various technological platforms (e.g. MS, NGS and HCS), a more in-depth view of cellular signaling can be obtained. This knowledge can subsequently be used for driving functional validation studies. We have demonstrated some conceptual and technological advancements which facilitate this process, ranging from how mutations should be interpreted functionally, through generating and modeling phospho-proteomic data, to integrating NGS and MS data to study the propagation of mutations at the protein level. Finally, we demonstrate the value of these approaches by attempting to apply them to the clinically severe problem of colon cancer metastasis. Due to the ubiquitous nature of phosphorylation based signaling in cellular decision making, we decided to focus mainly on this type of signaling, and thereby, the protein kinases, for which a great arsenal of available small molecule inhibitors and antibodies exists.

I predict that over the next years, MS and NGS technologies will become more prevalent in the clinic, as they have great potential in guiding therapeutic strategies in patients based on their tumors' protein signaling network states and genotypic profiles. Before this becomes feasible however, the large costs associated with these types of global screens will need to be significantly reduced. Additionally, while NGS allows the interrogation of the complete genome, Mass Spectrometry still needs to be improved in term of specificity. For example, while in Article 2 we describe how to generate phospho-proteomic data, it is uncertain where the upper limit lies in terms of numbers of phosphorylation sites which are being utilized by the cell. Given that KinomeXplorer-DB, a manually curated collection of published human phosphorylation sites based on Phospho.ELM[201] and PhosphoSitePlus[202], contains ~64,000 sites, our detection of ~30,000 in Article 5 is not even half way, and it is very likely that the total number of phosphorylation sites is much greater. Of course the phospho-proteome of a cell is dependent on cellular context, and our data only represents a snapshot of the phospho-proteome at a resting state while growing on collagen. Therefore, this data is still useful for modeling kinase-substrate interactions, as it highlights kinases which display enhanced activity under "normal" growing conditions, especially when used in combination with predictive algorithms such as KinomeXplorer. Advancements in this field are consistently on-going, resulting in greater coverage of cellular phospho-proteomes, and we are likely to see this pattern continue[203]. A current limitation of KinomeXplorer is that it does not possess comprehensive kinome-coverage in human, leading to a subset of phosphorylation sites to either be predicted for wrongly, or not at all, but large-scale efforts of closing this knowledge gap are currently on-going by systematically assessing the specificity of all kinases contained in the human kinome.

An improvement in sensitivity will also benefit our suggested approach of combining NGS and MS data, as this will lead to a greater number of mutations to be able to be monitored dynamically at the protein level. This will help start inferring function to certain mutations, as the regulation of them at the protein level may reveal whether they are involved in a specific disease phenotype or not. Alternatively, targeted MS using SRM[204,205] is likely to provide an advantage in this respect, as it allows the selective monitoring of mutated peptides in a complex sample.

Finally, a genome-wide assessment of cancer cell lines, including genome-wide RNAi screens are likely to reveal many more protein candidates with therapeutic potential, through which clinical benefit can possibly be obtained. We envision the approach we chose in Article 5 to be applicable to most cancer types, and additionally, different key cellular players in cancer development can likely undergo similar investigations. For example cancer-associated fibroblasts or cancer stem cells are likely to be defined by specific proteomic and genomic landscapes, and the elucidation thereof through an integrated approach should highlight key proteins which could be subjected to therapeutic interventions.

Bibliography

1. Society, A. C. The History of Cancer. (2012).
2. Hanahan, D. Rethinking the war on cancer. *Lancet* (2013).
3. Eckhouse, S., Lewison, G. & Sullivan, R. Trends in the global funding and activity of cancer research. *Mol Oncol* **2**, 20-32 (2008).
4. Gupta, G. P. & Massague, J. Cancer metastasis: building a framework. *Cell* **127**, 679-695 (2006).
5. Monge, J. et al. Fibrous dysplasia in a 120,000+ year old Neandertal from Krapina, Croatia. *PLoS One* **8**, e64539 (2013).
6. Stephens, P. J. et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400-404 (2012).
7. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546-1558 (2013).
8. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
9. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
10. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338-345 (2013).
11. Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328-337 (2013).
12. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat Med* **10**, 789-799 (2004).
13. Dawson, M. A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27 (2012).
14. Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nat Rev Cancer* **4**, 143-153 (2004).
15. Cox, J. & Mann, M. Is proteomics the new genomics? *Cell* **130**, 395-398 (2007).
16. Creixell, P., Schoof, E. M., Erler, J. T. & Linding, R. Navigating cancer network attractors for tumor-specific therapy. *Nat Biotechnol* **30**, 842-848 (2012).
17. Pawson, T. & Hunter, T. Signal transduction and growth control in normal and cancer cells. *Curr Opin Genet Dev* **4**, 1-4 (1994).
18. Pawson, T. & Kofler, M. Kinome signaling through regulated protein-protein interactions in normal and cancer cells. *Curr Opin Cell Biol* **21**, 147-153 (2009).
19. Tan, C. S. et al. Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci Signal* **2**, ra39 (2009).
20. Taylor, I. W. et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol* **27**, 199-204 (2009).
21. Taylor, I. W. & Wrana, J. L. Protein interaction networks in medicine and disease. *Proteomics* **12**, 1706-1716 (2012).
22. Vidal, M., Cusick, M. E. & Barabasi, A. L. Interactome networks and human disease. *Cell* **144**, 986-998 (2011).
23. Pawson, T. & Linding, R. Network medicine. *FEBS Lett* **582**, 1266-1270 (2008).
24. Davies, H. et al. Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-954 (2002).
25. Davies, M. A. & Samuels, Y. Analysis of the genome to personalize therapy for melanoma. *Oncogene* **29**, 5545-5555 (2010).
26. Flaherty, K. T. et al. Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med* **363**, 809-819 (2010).
27. Poulikakos, P. I. & Rosen, N. Mutant BRAF melanomas--dependence and resistance. *Cancer Cell* **19**, 11-15 (2011).
28. Wagle, N. et al. Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J Clin Oncol* **29**, 3085-3096 (2011).

29. Lito, P. et al. Relief of profound feedback inhibition of mitogenic signaling by RAF inhibitors attenuates their activity in BRAFV600E melanomas. *Cancer Cell* **22**, 668-682 (2012).
30. Yaffe, M. B. The scientific drunk and the lamppost: massive sequencing efforts in cancer discovery and treatment. *Sci Signal* **6**, pe13 (2013).
31. Clevers, H. The cancer stem cell: premises, promises and challenges. *Nat Med* **17**, 313-319 (2011).
32. Dick, J. E. Stem cell concepts renew cancer research. *Blood* **112**, 4793-4807 (2008).
33. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106-110 (2007).
34. Furth, J. & Kahn, M. The transmission of leukemia of mice with a single cell. *Am. J. Cancer* **31**, 276-282 (1937).
35. Bruce, W. R. & H, V. d. G. A QUANTITATIVE ASSAY FOR THE NUMBER OF MURINE LYMPHOMA CELLS CAPABLE OF PROLIFERATION IN VIVO. *Nature* **199**, 79-80 (1963).
36. Hewitt, H. B. Studies of the dissemination and quantitative transplantation of a lymphocytic leukaemia of CBA mice. *Br J Cancer* **12**, 378-401 (1958).
37. Makino, S. Further evidence favoring the concept of the stem cell in ascites tumors of rats. *Ann N Y Acad Sci* **63**, 818-830 (1956).
38. LJ, K. & Pierce, G. B. J. MULTIPOTENTIALITY OF SINGLE EMBRYONAL CARCINOMA CELLS. *Cancer Res* **24**, 1544-1551 (1964).
39. Lapidot, T. et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645-648 (1994).
40. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A* **100**, 3983-3988 (2003).
41. Singh, S. K. et al. Identification of human brain tumour initiating cells. *Nature* **432**, 396-401 (2004).
42. Bao, S. et al. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756-760 (2006).
43. Kreso, A. et al. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* **339**, 543-548 (2013).
44. Li, X. et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. *J Natl Cancer Inst* **100**, 672-679 (2008).
45. Zhou, B. B. et al. Tumour-initiating cells: challenges and opportunities for anticancer drug discovery. *Nat Rev Drug Discov* **8**, 806-823 (2009).
46. Kreso, A. et al. Self-renewal as a therapeutic target in human colorectal cancer. *Nat Med* **20**, 29-36 (2014).
47. O'Brien, C. A. et al. ID1 and ID3 regulate the self-renewal capacity of human colon cancer-initiating cells through p21. *Cancer Cell* **21**, 777-792 (2012).
48. Todaro, M. et al. Colon cancer stem cells dictate tumor growth and resist cell death by production of interleukin-4. *Cell Stem Cell* **1**, 389-402 (2007).
49. Dotto, G. P., Weinberg, R. A. & Ariza, A. Malignant transformation of mouse primary keratinocytes by Harvey sarcoma virus and its modulation by surrounding normal cells. *Proc Natl Acad Sci U S A* **85**, 6389-6393 (1988).
50. Orimo, A. et al. Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**, 335-348 (2005).
51. Polanska, U. M. & Orimo, A. Carcinoma-associated fibroblasts: non-neoplastic tumour-promoting mesenchymal cells. *J Cell Physiol* **228**, 1651-1657 (2013).
52. Bergers, G. et al. Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nat Cell Biol* **2**, 737-744 (2000).
53. Erler, J. T. et al. Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature* **440**, 1222-1226 (2006).

54. Fujita, H. et al. alpha-Smooth Muscle Actin Expressing Stroma Promotes an Aggressive Tumor Biology in Pancreatic Ductal Adenocarcinoma. *Pancreas* (2010).
55. Yamashita, M. et al. Role of stromal myofibroblasts in invasive breast cancer: stromal expression of alpha-smooth muscle actin correlates with worse clinical outcome. *Breast Cancer* **19**, 170-176 (2012).
56. Vihinen, P. & Kahari, V. M. Matrix metalloproteinases in cancer: prognostic markers and therapeutic targets. *Int J Cancer* **99**, 157-166 (2002).
57. Meads, M. B., Gatenby, R. A. & Dalton, W. S. Environment-mediated drug resistance: a major contributor to minimal residual disease. *Nat Rev Cancer* **9**, 665-674 (2009).
58. Straussman, R. et al. Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature* **487**, 500-504 (2012).
59. Wilson, T. R. et al. Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors. *Nature* **487**, 505-509 (2012).
60. Barker, H. E., Bird, D., Lang, G. & Erler, J. T. Tumor-secreted LOXL2 activates fibroblasts through FAK signaling. *Mol Cancer Res* **11**, 1425-1436 (2013).
61. Cox, T. R. et al. LOX-mediated collagen crosslinking is responsible for fibrosis-enhanced metastasis. *Cancer Res* **73**, 1721-1732 (2013).
62. Bissell, M. J. & Labarge, M. A. Context, tissue plasticity, and cancer: are tumor stem cells also regulated by the microenvironment? *Cancer Cell* **7**, 17-23 (2005).
63. Luga, V. et al. Exosomes mediate stromal mobilization of autocrine Wnt-PCP signaling in breast cancer cell migration. *Cell* **151**, 1542-1556 (2012).
64. Erler, J. T. & Giaccia, A. J. in *Clinical Oncology* 2008).
65. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* **2**, 563-572 (2002).
66. Fidler, I. J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat Rev Cancer* **3**, 453-458 (2003).
67. Hernando, E. et al. Rb inactivation promotes genomic instability by uncoupling cell cycle progression from mitotic control. *Nature* **430**, 797-802 (2004).
68. Maser, R. S. & DePinho, R. A. Connecting chromosomes, crisis, and cancer. *Science* **297**, 565-569 (2002).
69. Puc, J. et al. Lack of PTEN sequesters CHK1 and initiates genetic instability. *Cancer Cell* **7**, 193-204 (2005).
70. Gorgoulis, V. G. et al. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907-913 (2005).
71. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767 (1990).
72. Fidler, I. J. Metastasis: guantitative analysis of distribution and fate of tumor embolilabeled with 125 I-5-iodo-2'-deoxyuridine. *J Natl Cancer Inst* **45**, 773-782 (1970).
73. Kim, J. W. et al. Rapid apoptosis in the pulmonary vasculature distinguishes non-metastatic from metastatic melanoma cells. *Cancer Lett* **213**, 203-212 (2004).
74. Luzzi, K. J. et al. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am J Pathol* **153**, 865-873 (1998).
75. Paget, S. The distribution of secondary growths in cancer of the breast. *Lancet* **1**, 571-573 (1889).
76. Nguyen, D. X., Bos, P. D. & Massague, J. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer* **9**, 274-284 (2009).
77. Chang, J. & Erler, J. Hypoxia-mediated metastasis. *Adv Exp Med Biol* **772**, 55-81 (2014).
78. Harris, A. L. Hypoxia--a key regulatory factor in tumour growth. *Nat Rev Cancer* **2**, 38-47 (2002).

79. Staller, P. et al. Chemokine receptor CXCR4 downregulated by von Hippel-Lindau tumour suppressor pVHL. *Nature* **425**, 307-311 (2003).
80. Pennacchietti, S. et al. Hypoxia promotes invasive growth by transcriptional activation of the met protooncogene. *Cancer Cell* **3**, 347-361 (2003).
81. Hussain, S. P., Hofseth, L. J. & Harris, C. C. Radical causes of cancer. *Nat Rev Cancer* **3**, 276-285 (2003).
82. Paszek, M. J. et al. Tensional homeostasis and the malignant phenotype. *Cancer Cell* **8**, 241-254 (2005).
83. Baker, A. M., Bird, D., Lang, G., Cox, T. R. & Erler, J. T. Lysyl oxidase enzymatic function increases stiffness to drive colorectal cancer progression through FAK. *Oncogene* **32**, 1863-1868 (2013).
84. Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559-1564 (2011).
85. Brognard, J. & Hunter, T. Protein kinase signaling networks in cancer. *Curr Opin Genet Dev* **21**, 4-11 (2011).
86. Criscuoli, M. L., Nguyen, M. & Eliceiri, B. P. Tumor metastasis but not tumor growth is dependent on Src-mediated vascular permeability. *Blood* **105**, 1508-1514 (2005).
87. Engelman, J. A. et al. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316**, 1039-1043 (2007).
88. Hiratsuka, S. et al. MMP9 induction by vascular endothelial growth factor receptor-1 is involved in lung-specific metastasis. *Cancer Cell* **2**, 289-300 (2002).
89. Lee, M. J. et al. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780-794 (2012).
90. Regan Anderson, T. M. et al. Breast tumor kinase (Brk/PTK6) is a mediator of hypoxia-associated breast cancer progression. *Cancer Res* **73**, 5810-5820 (2013).
91. Jorgensen, C. et al. Cell-specific information processing in segregating populations of Eph receptor ephrin-expressing cells. *Science* **326**, 1502-1509 (2009).
92. Davis, M. I. et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* **29**, 1046-1051 (2011).
93. Fedorov, O., Muller, S. & Knapp, S. The (un)targeted cancer kinome. *Nat Chem Biol* **6**, 166-169 (2010).
94. Janne, P. A., Gray, N. & Settleman, J. Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. *Nat Rev Drug Discov* **8**, 709-723 (2009).
95. Karaman, M. W. et al. A quantitative analysis of kinase inhibitor selectivity. *Nat Biotechnol* **26**, 127-132 (2008).
96. Lin, J. et al. A multidimensional analysis of genes mutated in breast and colorectal cancers. *Genome Res* **17**, 1304-1318 (2007).
97. Wood, L. D. et al. The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108-1113 (2007).
98. Linding, R. et al. Systematic discovery of in vivo phosphorylation networks. *Cell* **129**, 1415-1426 (2007).
99. Macurek, L. et al. Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery. *Nature* **455**, 119-123 (2008).
100. Yaffe, M. B. & Cantley, L. C. Signal transduction. Grabbing phosphoproteins. *Nature* **402**, 30-31 (1999).
101. Jemal, A. et al. Cancer statistics, 2008. *CA Cancer J Clin* **58**, 71-96 (2008).
102. Markowitz, S. D., Dawson, D. M., Willis, J. & Willson, J. K. Focus on colon cancer. *Cancer Cell* **1**, 233-236 (2002).

103. Andre, T. et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med* **350**, 2343-2351 (2004).
104. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623-627 (1997).
105. Markowitz, S. D. & Bertagnolli, M. M. Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med* **361**, 2449-2460 (2009).
106. Al-Tassan, N. et al. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nat Genet* **30**, 227-232 (2002).
107. Bronner, C. E. et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368**, 258-261 (1994).
108. Herman, J. G. et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A* **95**, 6870-6875 (1998).
109. Kane, M. F. et al. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res* **57**, 808-811 (1997).
110. Kastrinos, F. & Syngal, S. Recently identified colon cancer predispositions: MYH and MSH6 mutations. *Semin Oncol* **34**, 418-424 (2007).
111. Leach, F. S. et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215-1225 (1993).
112. Bienz, M. & Clevers, H. Linking colorectal cancer to Wnt signaling. *Cell* **103**, 311-320 (2000).
113. Goss, K. H. & Groden, J. Biology of the adenomatous polyposis coli tumor suppressor. *J Clin Oncol* **18**, 1967-1979 (2000).
114. Sansom, O. J. et al. Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev* **18**, 1385-1390 (2004).
115. Wilkins, J. A. & Sansom, O. J. C-Myc is a critical mediator of the phenotypes of Apc loss in the intestine. *Cancer Res* **68**, 4963-4966 (2008).
116. Athineos, D. & Sansom, O. J. Myc heterozygosity attenuates the phenotypes of APC deficiency in the small intestine. *Oncogene* **29**, 2585-2590 (2010).
117. Korinek, V. et al. Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science* **275**, 1784-1787 (1997).
118. Morin, P. J. et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* **275**, 1787-1790 (1997).
119. Baker, S. J. et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**, 217-221 (1989).
120. Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K. & Vogelstein, B. Suppression of human colorectal carcinoma cell growth by wild-type p53. *Science* **249**, 912-915 (1990).
121. Baker, S. J. et al. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res* **50**, 7717-7722 (1990).
122. Vazquez, A., Bond, E. E., Levine, A. J. & Bond, G. L. The genetics of the p53 pathway, apoptosis and cancer therapy. *Nat Rev Drug Discov* **7**, 979-987 (2008).
123. Baker, S. J. et al. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res* **50**, 7717-7722 (1990).
124. Markowitz, S. et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268**, 1336-1338 (1995).
125. Eppert, K. et al. MADR2 maps to 18q21 and encodes a TGFbeta-regulated MAD-related protein that is functionally mutated in colorectal carcinoma. *Cell* **86**, 543-552 (1996).
126. Grady, W. M. et al. Mutational inactivation of transforming growth factor beta receptor type II in microsatellite stable colon cancers. *Cancer Res* **59**, 320-324 (1999).

127. Leary, R. J. et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc Natl Acad Sci U S A* **105**, 16224-16229 (2008).
128. Sjoblom, T. et al. The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-274 (2006).
129. Thiagalingam, S. et al. Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. *Nat Genet* **13**, 343-346 (1996).
130. Bos, J. L. et al. Prevalence of ras gene mutations in human colorectal cancers. *Nature* **327**, 293-297 (1987).
131. Rajagopalan, H. et al. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature* **418**, 934 (2002).
132. Siena, S., Sartore-Bianchi, A., Di Nicolantonio, F., Balfour, J. & Bardelli, A. Biomarkers predicting clinical outcome of epidermal growth factor receptor-targeted therapy in metastatic colorectal cancer. *J Natl Cancer Inst* **101**, 1308-1324 (2009).
133. Parsons, D. W. et al. Colorectal cancer: mutations in a signalling pathway. *Nature* **436**, 792 (2005).
134. Samuels, Y. et al. High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554 (2004).
135. Moertel, C. G. et al. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. *N Engl J Med* **322**, 352-358 (1990).
136. Thirion, P. et al. Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: an updated meta-analysis. *J Clin Oncol* **22**, 3766-3775 (2004).
137. Andre, T. et al. Improved overall survival with oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. *J Clin Oncol* **27**, 3109-3116 (2009).
138. Cunningham, D. & Starling, N. Adjuvant chemotherapy of colorectal cancer. *Lancet* **370**, 1980-1981 (2007).
139. Douillard, J. Y. et al. Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. *Lancet* **355**, 1041-1047 (2000).
140. Kohne, C. H. et al. Phase III study of weekly high-dose infusional fluorouracil plus folinic acid with or without irinotecan in patients with metastatic colorectal cancer: European Organisation for Research and Treatment of Cancer Gastrointestinal Group Study 40986. *J Clin Oncol* **23**, 4856-4865 (2005).
141. Kuebler, J. P. et al. Oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: results from NSABP C-07. *J Clin Oncol* **25**, 2198-2204 (2007).
142. Colucci, G. et al. Phase III randomized trial of FOLFIRI versus FOLFOX4 in the treatment of advanced colorectal cancer: a multicenter study of the Gruppo Oncologico Dell'Italia Meridionale. *J Clin Oncol* **23**, 4866-4875 (2005).
143. Allegra, C. J. et al. Phase III trial assessing bevacizumab in stages II and III carcinoma of the colon: results of NSABP protocol C-08. *J Clin Oncol* **29**, 11-16 (2011).
144. Bokemeyer, C. et al. Fluorouracil, leucovorin, and oxaliplatin with and without cetuximab in the first-line treatment of metastatic colorectal cancer. *J Clin Oncol* **27**, 663-671 (2009).
145. Gibson, T. B., Ranganathan, A. & Grothey, A. Randomized phase III trial results of panitumumab, a fully human anti-epidermal growth factor receptor monoclonal antibody, in metastatic colorectal cancer. *Clin Colorectal Cancer* **6**, 29-31 (2006).
146. Saltz, L. B. et al. Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. *J Clin Oncol* **26**, 2013-2019 (2008).
147. Van Cutsem, E. et al. Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med* **360**, 1408-1417 (2009).

148. Beadle, G. W. & Tatum, E. L. Genetic Control of Biochemical Reactions in Neurospora. *Proc Natl Acad Sci U S A* **27**, 499-506 (1941).
149. Chandarlapaty, S. et al. AKT inhibition relieves feedback suppression of receptor tyrosine kinase expression and activity. *Cancer Cell* **19**, 58-71 (2011).
150. Huang, P. H. et al. Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc Natl Acad Sci U S A* **104**, 12867-12872 (2007).
151. Schoeberl, B. et al. Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. *Sci Signal* **2**, ra31 (2009).
152. Gstaiger, M. & Aebersold, R. Genotype-phenotype relationships in light of a modular protein interaction landscape. *Mol Biosyst* **9**, 1064-1067 (2013).
153. Bakal, C., Aach, J., Church, G. & Perrimon, N. Quantitative morphological signatures define local signaling networks regulating cell morphology. *Science* **316**, 1753-1756 (2007).
154. Nurse, P. The great ideas of biology. *Clin Med* **3**, 560-568 (2003).
155. Vidal, M. A unifying view of 21st century systems biology. *FEBS Lett* **583**, 3891-3894 (2009).
156. Gstaiger, M. & Aebersold, R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet* **10**, 617-627 (2009).
157. Miller-Jensen, K., Janes, K. A., Brugge, J. S. & Lauffenburger, D. A. Common effector processing mediates cell-specific responses to stimuli. *Nature* **448**, 604-608 (2007).
158. Huang, P. H., Cavenee, W. K., Furnari, F. B. & White, F. M. Uncovering therapeutic targets for glioblastoma: a systems biology approach. *Cell Cycle* **6**, 2750-2754 (2007).
159. Janes, K. A. et al. A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* **310**, 1646-1653 (2005).
160. Jorgensen, C. & Linding, R. Directional and quantitative phosphorylation networks. *Brief Funct Genomic Proteomic* **7**, 17-26 (2008).
161. Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117-1134 (2010).
162. Linding, R. et al. NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res* **36**, D695-9 (2008).
163. Miller, M. L. et al. Linear motif atlas for phosphorylation-dependent signaling. *Sci Signal* **1**, ra2 (2008).
164. Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633-1649 (2004).
165. Obenauer, J. C., Cantley, L. C. & Yaffe, M. B. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* **31**, 3635-3641 (2003).
166. Wong, Y. H. et al. KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res* **35**, W588-94 (2007).
167. Xue, Y. et al. GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res* **33**, W184-7 (2005).
168. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719-724 (2009).
169. Network, T. C. G. A. R. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609-615 (2011).
170. Forbes, S. A. et al. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res* **38**, D652-7 (2010).
171. Pleasance, E. D. et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184-190 (2010).

172. Rapley, E. A. et al. A genome-wide association study of testicular germ cell tumor. *Nat Genet* **41**, 807-810 (2009).
173. Stephens, P. J. et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**, 1005-1010 (2009).
174. Torkamani, A. & Schork, N. J. Identification of rare cancer driver mutations by network reconstruction. *Genome Res* **19**, 1570-1578 (2009).
175. Beck, M. et al. The quantitative proteome of a human cell line. *Mol Syst Biol* **7**, 549 (2011).
176. Geiger, T., Wehner, A., Schaab, C., Cox, J. & Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics* **11**, M111.014050 (2012).
177. Munoz, J. & Heck, A. J. Quantitative proteome and phosphoproteome analysis of human pluripotent stem cells. *Methods Mol Biol* **767**, 297-312 (2011).
178. Nagaraj, N. et al. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* **7**, 548 (2011).
179. Wisniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).
180. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B. & Aebersold, R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods* **4**, 231-237 (2007).
181. Bodenmiller, B. & Aebersold, R. Quantitative analysis of protein phosphorylation on a system-wide scale by mass spectrometry-based proteomics. *Methods Enzymol* **470**, 317-334 (2010).
182. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* **4**, 484-494 (2009).
183. Ong, S. E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).
184. Geiger, T. et al. Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat Protoc* **6**, 147-157 (2011).
185. Echeverri, C. J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* **7**, 373-384 (2006).
186. Mohr, S., Bakal, C. & Perrimon, N. Genomic screening with RNAi: results and challenges. *Annu Rev Biochem* **79**, 37-64 (2010).
187. Mohr, S. E. & Perrimon, N. RNAi screening: new approaches, understandings, and organisms. *Wiley Interdiscip Rev RNA* **3**, 145-158 (2012).
188. Moffat, J. & Sabatini, D. M. Building mammalian signalling pathways with RNAi screens. *Nat Rev Mol Cell Biol* **7**, 177-187 (2006).
189. Dorsett, Y. & Tuschl, T. siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov* **3**, 318-329 (2004).
190. Wilson, J. L., Hemann, M. T., Fraenkel, E. & Lauffenburger, D. A. Integrated network analyses for functional genomic studies in cancer. *Semin Cancer Biol* **23**, 213-218 (2013).
191. Weiss, W. A., Taylor, S. S. & Shokat, K. M. Recognizing and exploiting differences between RNAi and small-molecule inhibitors. *Nat Chem Biol* **3**, 739-744 (2007).
192. Azorsa, D. O. et al. Synthetic lethal RNAi screening identifies sensitizing targets for gemcitabine therapy in pancreatic cancer. *J Transl Med* **7**, 43 (2009).
193. Kaelin, W. G. J. The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer* **5**, 689-698 (2005).
194. Luo, J. et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835-848 (2009).
195. Vizeacoumar, F. J. et al. A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. *Mol Syst Biol* **9**, 696 (2013).

196. Kim, H. D. et al. Signaling network state predicts twist-mediated effects on breast cell migration across diverse growth factor contexts. *Mol Cell Proteomics* **10**, M111.008433 (2011).
197. Kreeger, P. K. & Lauffenburger, D. A. Cancer systems biology: a network modeling perspective. *Carcinogenesis* **31**, 2-8 (2010).
198. Morris, M. K., Saez-Rodriguez, J., Sorger, P. K. & Lauffenburger, D. A. Logic-based models for the analysis of cell signaling networks. *Biochemistry* **49**, 3216-3224 (2010).
199. Pritchard, J. R., Bruno, P. M., Hemann, M. T. & Lauffenburger, D. A. Predicting cancer drug mechanisms of action using molecular network signatures. *Mol Biosyst* **9**, 1604-1619 (2013).
200. Saez-Rodriguez, J. et al. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol* **5**, 331 (2009).
201. Dinkel, H. et al. Phospho.ELM: a database of phosphorylation sites--update 2011. *Nucleic Acids Res* **39**, D261-7 (2011).
202. Hornbeck, P. V. et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* **40**, D261-70 (2012).
203. Zhou, H. et al. Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat Protoc* **8**, 461-480 (2013).
204. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **138**, 795-806 (2009).
205. Picotti, P., Bodenmiller, B. & Aebersold, R. Proteomics meets the scientific method. *Nat Methods* **10**, 24-27 (2013).