Proteomic approaches for quantitative cancer cell signaling

Voellmy, Franziska; Sommer, Morten Otto Alexander; Linding, Rune

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Doctoral Thesis

Proteomic approaches for quantitative cancer cell signaling

Author:
Franziska Voellmy

Supervisors:
Morten Sommer
Rune Linding

July 27, 2017

Preface

This thesis is submitted as a requirement for obtaining the Ph.D. degree at the Technical University of Denmark (DTU), and was funded by a DTU scholarship. The work was conducted under the joint supervision of Prof Morten Sommer at the Novo Nordisk Foundation Center for Biosustainability (CfB), and Prof Rune Linding at the Biotech Research & Innovation Centre (BRIC) at the University of Copenhagen (KU). The project was formally hosted by the CfB under the supervision of Morten Sommer.

Lyngby, July 2017
Franziska Voellmy

Abstract

Cancer is a genetic disease and historically the discovery of underlying genetic alterations has been critical to our understanding of disease and past treatment successes. However, cancer still poses an important health threat and most available drugs are not capable of providing complete remission. Drug targets are typically proteins, but are based on genetic findings. Thus studying cancer systems at the level of proteins and their signaling can provide the additional level of data needed for the development of effective drugs. This thesis summarizes the work undertaken during my doctoral studies in an effort to contribute to the study of signaling dynamics in cancer systems.

This thesis is divided into two parts. Part I begins with a brief introduction in the use of omics in systems cancer research with a focus on mass spectrometry as a means to quantitatively measure protein and signaling dynamics in the identified protein networks (Chapter 1). Gene fusions are portrayed in-depth as an example of a major genetic alteration found to occur in a variety of cancers, the most infamous of which has lead to the development of the specific tyrosine kinase inhibitor imatinib and a major success in the treatment of chronic myelogenous leukemia. However, this is the exception rather than the norm as most drugs are developed based on genetic findings while designed to act on the protein level, and might contribute to explaining the paucity of specific effective cancer therapeutics available. Furthermore, this underlines the importance of proteomic studies and the conclusions drawn for the high-throughput data generated in the latter. Chapter 2 gives a temporal overview of precision gene-editing in the context of systems biology. Following the past successes of methods such as zinc-finger nucleases and TALEs, the novel CRISPR-Cas technology has rapidly become an extremely popular gene-editing tool. Its mechanism of action, several applications and potential shortcomings are discussed. The Chapter is concluded with a final application: chromosomal translocations can be generated *in vitro* or *in vivo* using nuclease-based targeted gene-editing. Part II illustrates the use of mass spectrometry-based proteomics and phospho-proteomics in studying the effects of perturbations at the cellular level. Chapter three captures the very early signaling dynamics related to cell migration following wounding in triple negative breast cancer cells, and their potential role as novel targets for therapies aimed at reducing the metastases. Chapter four describes the induction of the oncogenic chimeric gene PRKAR1A-RET in thyroid cells. Its transformative potential is shown and the ensuing changes are measured at the protein and signaling levels. This study demonstrates the use of the novel CRISPR-Cas technology for the generation of chromosomal rearrangements *in vitro* and investigates the effects of this important genetic aberration in a physiologically relevant cellular setting. Part III concludes the thesis by providing a global discussion and future perspectives for the studies presented in part II.

Overall, the work presented herein aims to underscore the importance of studying cancer systems at the protein level, the dynamics of which define phenotypic outcome. The effects of cellular and genetic perturbations at the protein network level were studied using mass spectrometry-based proteomics

and, the results whereof suggest interesting avenues for future development of cancer therapies.

Dansk Resume

Cancer er en genetisk sygdom og historisk set har kortlægningen af de underliggende genetiske ændringer været central for vores forståelse af sygdommen og grundlaget for tidligere behandlingsformer's succes. Alligevel er cancer til stadighed en stor sundhedstrussel. De fleste behandlingsstrategier er ikke i stand til fuldt at helbrede patienter. Medicinsk behandling er typisk målrettet proteiner, men forskningen bag dem er baseret pågenetiske opdagelser. Et nyt domæne af viden kan bidrage til ny effektiv medicin ved direkte at studere cancersystemer med udgangspunkt i proteiner og deres signalering. denne afhandling opsummerer mit Ph.d. arbejde, og har til hensigt at bidrage til forståelsen af signalering i dynamiske cancersystemer.

denne afhandling er opdelt i to. Del I begynder med en kort introduktion til anvendelsen af 'omics' i systembiologisk cancerforskning. Her vil jeg fokusere påmassespektroskopi som kvantitativ målemetode af proteiner og signaleringsdynamikker i identificerede protein netværk (kapitel 1). Genfusioner bliver grundigt beskrevet, som et eksempel påen voldsom genetisk modifikation. De findes i en lang række af cancertyper, hvor iblandt opdagelsen af den mest kendte genfusion har ført til udviklingen en specifik tyrosin kinase inhibitor, Imatinib, og herfra en stor forbedring i behandlingen af kronisk myeloid leukæmi. Imatinib er dog undtagelsen der bekræfter reglen, da det meste medicin er udviklet fra genetisk viden, men designet til at virke pået proteinniveau, hvilket muligvis forklarer den nuværende mangel påeffektive cancer behandlingsformer. Dette understreger betydningen af proteomiske studier og de konklusioner der drages fra den store mængde viden der generes i disse. Kaptiel to giver et overblik i tid, af præcis genmanipulering i den systembiologiske kontekst. I kølvandet påtidligere succesfulde teknikker, som zinc-finger nukleaser og TALEs, er den nye CRISPR-Cas teknologi hurtigt blevet en ekstremt populært værktøj til genmodificering. Dets virkningsmekanisme, forskellige anvendelser og potentielle begrænsninger beskrives. Kapitlet afrundes med en anvendelse: kromosomal translokationer kan genereres *in vitro* eller *in vivo* vha. målrettede nukleasegenmanipulering. Del II illustrerer brugen af massespektroskopi og phosphoproteomik i studiet af perturbationer pådet cellulære niveau. Kapitel tre beskrives de meget tidlige signal-dynamikker relateret til migration som følge af sår-ridsning i triple negativ brystkræft, og deres potentielle rolle som nye mål for behandlinger der er skal reducerer metastaser. Kapitel fire beskriver det onkogenske kimære gen PRKAR1A-RET i celler fra skjoldbruskkirtlen. Dets transformative potentiale vises og de efterfølgende virkning påprotein- og signaleringsniveau måles. Dette studie demonstrerer brugen af den nye CRISPR-Cas teknologi i skabelsen af kromosomale omrokeringer *in vitro* og undersøger effekten af denne betydningsfulde genetiske afvigelse i en fysiologisk relevant cellulær kontekst. Del III opsamler specialet med en generel diskussion og præsenterer fremtidige perspektiver for arbejdet i del II.

Alt i alt er målet med det fremlagte arbejde at understrege vigtigheden af at studere cancersystemer fra et protein-perspektiv, hvis dynamikker bestemmer det fænotypeiske udtryk. Konsekvensen af cellulære og genetiske

perturbationer påcellens proteiner blev studeret med proteomik baseret på-massespektroskopi og resultaterne herfra antyder nye interessante muligheder for fremtidig udvikling af behandlingsstrategier mod cancer.

Acknowledgements

The work laid out in this thesis would not have been possible without the support and encouragement of many people whom I'd like to thank here. I would first like to express my gratitude to my supervisors Rune Linding and Morten Sommer. Rune, your passion and enthusiasm for research is what triggered my move to Denmark and these qualities have been a constant source of inspiration throughout my PhD. Together with Morten, you have created an environment of scientific freedom that I was fortunate enough to benefit from. This has been incredibly motivating and has and the confidence you have had in me since the beginning has allowed to me to massively expand my skill set.

I cannot thank James Longden enough, who on many occasions took the role of my supervisor, providing guidance and support on an everyday basis. Without a doubt, none of this would have been possible without you. Thank you for indulging me and my usual scientific grumpiness, and for being incredibly tolerant. After discussions, my confidence and motivation to do yet another experiment is renewed :)

An immense thanks then goes to the past and present members of the Linding lab who have all been incredible colleagues. I have learnt so much from each of you over the years. Creativity, scientific rigor, tolerance, patience and enthusiasm are just some of the qualities I continue to be inspired by. Specifically, thanks to Craig for being my North American buddy :) and a mentor to me in terms of experimental works as well as project design. Perhaps all my questions in Colorado were annoying, but I greatly enjoyed the discussions that shaped my main project as it is now! Speaking of, I'd also like to thank Conor Howard, who spent some time in the lab as a student and helped me out a lot with cloning. Xavier, you were essential to all of the projects I participated during my PhD, and your patience makes you an awesome teacher for all things analysis coding-related. To Gaye, a huge thanks for many discussions, your advice, and being such a fun person to be around. Rock on, girl ! Mathias, I think we're going to have a lot of fun with all the data ! Also, thank you for the help with the Danish translation. Thanks to Tom Altenburg for your input for several projects despite your short time in the lab, hope to see you soon in the lab. Thanks go out as well to former lab members: Erwin, Cristina, Pau, Oxana, Jinho and Jesper. Our time together was shorter but I definitely learnt a lot and your contributions have made the lab what it is. Especially Erwin, who gave me a MS crash course when I started out as a student in the lab. Additionally, I'm thankful for the collaborations with the Erler lab at BRIC. In particular, Tom, Alejandro and Anette were awesome to work alongside. I would also like to thank all those who have agreed to comment on my thesis, especially when time was of the essence. This includes the members of the Linding lab as well as my friend Emma Hagberg.

I would also like to thank some members of ThermoFisher's scientific team for great support for any LC-MS/MS-related issues throughout the years, and I have learnt so much about mass spectrometry (and its troubleshooting) thanks to them. In no particular order, thanks go to : Lars H., Lars K., Erik, Jess, Mikael, Romain, and Vlad.

My friends and partner have in so many ways contributed to making this journey a fun one ! To my closest Swiss friends Rosalie, Marco, Jane: I am

forever grateful for your support despite the distance and for your efforts to maintain the contact and visiting me here in Copenhagen. To Anet, I am so glad that we made the move to Denmark together. It was awesome to have a great friend here from the beginning! A shoutout also goes to my good friends Diane and Emma in Copenhagen. All those coffees and dinners were necessary for my survival here :) A special thanks also goes to Frank. With your support and great sense of humour, you have helped turn the challenges associated with this thesis into a fun project: the Flapthesis ;D

I would like to thank my family who have been supportive of this entire journey. To my brothers Lukas and Andreas for math and biology discussions on the ski lift; to my mother Alice for setting an example of confidence and never-ending energy. Finally, I dedicate this work to my father, Richard, who has been an inspiration to me in terms of science and critical thinking, and has always encouraging me to pursue my interest in research.

Papers included in the thesis

Voellmy, F. and Linding, R (2017) "Precision Genome Editing for Systems Biology - A Temporal Perspective", in Systems Biology (eds J. Nielsen and S. Hohmann), Wiley-VCH Verlag GmbH &Co. KGaA, Weinheim, Germany.

Voellmy, F., Simpson, C., Howard, C., Saginc, G., Robin, X., Longden, J., Linding, R. "Exploring the PRKAR1A-RET induced oncogenic transformation of thyroid epithelial cells".

Voellmy, F., Engel, M., Saginc, G., Robin, X., Longden, J., Linding, R. "Proteogenomic analysis of dynamic cell signaling responses to wounding".

Papers not included in the thesis

Simon Koplev, James Longden, Jesper Ferkinghoff-Borg, Mathias Blicher Bjerregård, Thomas R. Cox, Jesper T. Pedersen, **Franziska Voellmy**, Janine T. Erler, Morten O.A. Sommer and Rune Linding, "Dynamic rearrangement of cell states detected by systematic screening of sequential anticancer treatments". *In submission at Cell Reports.*

Xavier Robin, Jesper Ferkinghoff-Borg, Erwin Schoofe, **Franziska Voellmy**, Jinho Kim, James Longden, Graham Murray, Rune Linding. "Integrative modeling of proteomic and genomic data predicts metastatic signatures in colorectal cancer patients". *Manuscript in preparation.*

Ralitsa R. Madsen, Rachel G. Knox, **Franziska Voellmy**, Xavier Robin, James Longden, Rune Linding, Robert K. Semple. "Signalling dynamics in human pluripotent stem cells with activating PIK3CA mutations." *To be submitted to Cell Stem Cell.*

Laura Arribas-Hernandez, Anja Branscheid, Axel Thieffry, Erwin Schoof, **Franziska Voellmy**, Rune Linding, Peter Brodersen. "Physical association of exosome co-factors with RISC limits secondary siRNA formation and explains the requirement for 2'-O-methylation of plant miRNAs" *To be submitted to Genes&Development.*

Lars Sjögren, Maïna Floris, Andrea Barghetti, **Franziska Voellmy**, Rune Linding, Peter Brodersen. "Farnesylated Heat Shock Protein 40 associates with membrane-bound ARGONAUTE1 in Arabidopsis." *To be submitted to Plant Cell.*

Majorca, A., Robin, X., **Voellmy F**, Jesper Ferkinghoff-Borg, Longden, J., Erler, J., Linding, R. "Homing mechanism identifies signatures of sites of metastasis in mouse". *Manuscript in preparation*

Contents

List of Figures

List of Supplementary Figures

List of Tables

List of Abbreviations

MS	**M**ass **S**pectrometry
LC-MS/MS	Tandem **L**iquid **C**hromatography **M**ass spectrometry
SNP	**S**ingle **N**ucleotide **P**olymorphism
GWA	**G**enome-**W**ide **A**ssociation
GWAS	**G**enome-**W**ide **A**ssociation **S**tudy
WHO	**W**orld **H**ealth **O**rganization
ChIP-seq	**Ch**romatin **I**mmuno**P**recipitation **seq**uencing
RNA-seq	**R**ibo**N**ucleic **A**cid **seq**uencing
CML	**C**hronic **M**yelogenous **L**eukemia
Ser	**Ser**ine
Thr	**Thr**eonine
Tyr	**Tyr**osine
GO	**G**ene **O**ntology
ECM	**E**xtracellular **M**atrix
TCGA	**T**he **C**ancer **G**enome **A**tlas
TALEN	**T**ranscription-**A**ctivator-**L**ike **E**ffector **N**ucleases
ZFN	**Z**inc **F**inger **N**uclease
CRISPR	**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats
DSB	**D**ouble-**S**trand **B**reak
PAM	**P**rotospacer **A**djacent **M**otif
DBD	**D**NA-**B**inding **D**omain
HDR	**H**omology-**D**irected **R**epair
NHEJ	**N**on-**H**omologous **E**nd **J**oining
TK	**T**yrosine **K**inase
RTK	**R**eceptor **T**yrosine **K**inase
FTK	**F**usion **T**yrosine **K**inase

Part I

Introduction

The first part of this thesis aims to provide a brief introduction to the use of high-throughput technologies in cancer research since the advent of sequencing, with a focus on genomics and proteomics. These technologies are characterized by their ability to generate data in a high-throughput fashion and as a consequence of the breadth of data generated, allow global, systematic studies that were previously difficult to conduct. After a succint overview of these large-scale studies within cancer research, two different ways of perturbing the cellular protein network are outlined. The first is by creating a direct perturbation at the level of the cell's signaling network. The second is a genetic pertubation. Indeed, chromosomal rearrangements are introduced as an example of the genetic basis of cancer for which targeted therapy exists in the form of a kinase inhibitor, as it around these that a large part of the work carried out in this thesis is centered. The nature and importance of chromosomal rearrangements, a genetic perturbation to the cell, are discussed in detail. Furthermore, the current standard workflow in terms of integration of data sets with the hope of developing novel therapeutic strategies is outlined. Building on the topic of chromosomal rearrangements, chapter 2 is a published book chapter which provides a historical perspective on gene editing and the popular CRISPR-Cas system, setting the stage for the use of this powerful technology to generate and study gene fusions.

Chapter 1

Investigating cancer signaling using mass spectrometry-based proteomics

1.1 Cancer

Despite ancient civilizations already being confronted with cancer [1], we continue to struggle to understand and most importantly, to develop effective means to treat it. The World Health Organization (WHO) reported cancer to be the second leading cause of death worldwide, with a death toll of 8.8 million in 2015 [2]. The incidence of cancer continues to rise, being the highest in high-income countries in Western Europe, North America, Japan, the Republic of Korea, Australia and New Zealand. This can be attributed to advances in general medicine which have led to longer life expectancy, thus resulting in a higher number of incident cases due to cancer's strong relationship with age. This improvement in longevity in addition to increased exposure to known carcinogens such as tobacco or UV radiation has led to cancer topping the list of causes of mortality globally.

1.1.1 Cancer: a disease of networks

Major biological events such as cell division or migration involve both intra- and inter-cellular interactions and the organization of these networks of cells paints a complex picture of synergy between various cellular components, such as genes, proteins, and metabolites. Indeed, signaling networks have evolved such that cells can integrate multiple input cues in order to collectively respond to changes in the environment. With this in mind, the central dogma of biology describing the flow of genetic information from DNA to proteins via messenger RNA can no longer be viewed in a strictly linear fashion nor at the level of a single gene [3]. Instead the large number of cellular players should be considered as constituting a connected network and participating in dynamic and non-linear processes, with features such as feedback and crosstalk [4].

The undertaking of the human genome project in the '90s helped bring about the latter holistic description of biological systems. It also spurred the development of novel high-throughput technologies such as mass spectrometry applied to the large-scale study of proteins, which will be discussed in section 1.4.2. The field of systems biology employs these technologies and develops models aiming to describe how the parts of a network give rise to

the behavior of the system as a whole [4]. A biological system generally exhibits emergent features that would not be apparent when investigating its individual parts. The former "reductionist" approach has successfully inventoried the majority of the cellular players and many of their interactions, but has failed to provide a working model of how properties specific to systems arise. With the continued improvement of instrumentation and development of novel technologies, large lists of genes, transcripts, metabolites, proteins and more have been generated, and the focus has shifted toward a global quantitative view of the biological system at hand, rather than an isolated gene or pathway [5]. These technologies have been used to better understand events occurring within the cell, such as determining breast cancer biomarkers from system-wide protein-protein interaction data [6].

This approach has given weight to the belief that in order to understand complex biological systems, questions require a multi-pronged approach with data representing the multiple layers of complexity. Given the biological layers of DNA transcription into RNA, followed by mRNA translation leading to protein expression, and finally post-translational modification of proteins, technologies such as genomics, transcriptomics, metabolomics, epigenetics and proteomics, as well as phenotypic data from imaging should be used in concert and modeled to produce a more complete snapshot of the process of interest. The emergence of technologies capable of acquiring data at a high throughput such as deep sequencing and quantitative mass spectrometry-based proteomics enable these approaches.

Uncontrolled proliferation and metastases are just two of the most prominent characteristics of cancerous cells that have been discovered by scientific advances in recent decades. Progress in the clinic has been furthered with the development of chemotherapeutics to treat patients suffering from cancer or to alleviate symptoms and thereby improve life expectancy. Drugs developed to address these characteristics typically possess a substantial level of toxicity for the patient, since the uncontrolled processes are relevant in healthy individuals and it is thus difficult to target only the malignant cells. More recently, the genomic revolution (and the sequencing of the full human genome [7]–[9] triggered by easier access to DNA sequencing technology as well as reduced cost has enabled many studies aiming to identify the major genetic players in the onset and progression of cancer.

As a result of a large number of sequencing studies (described in more detail in section 1.1.2), countless genomic modifications such as single-nucleotide polymorphisms, deletions and frameshifts are being identified in cancer genomes. In fact, chromosomal aberrations such as insertions, deletions, or translocations require genetic technologies in order to be identified during patient tumor profiling.

While cancer has historically been considered a genetic disease and the detection of genetic aberrations is crucial to our understanding of disease, these alterations alone do not provide complete insight into the question of how cells become cancerous. Indeed, cancer has emerged as a multi-faceted disease with genetic mutations, regulation by epigenetic markers [10], and dysregulation of core cellular processes such as proliferation, migration and homeostasis relying on signaling and interactions at the protein level [11]. Due to underlying genetic alterations cancer cells respond differently to the input cues received from the environment compared to healthy cells. This allows malignant cells to bypass the normally tightly regulated collective cell

population dynamics and instead make decisions purely for their own benefit, resulting for instance in uncontrolled proliferation.

In an attempt to rationalize the complex phenomena underlying tumor growth and metastatic dissemination, Hanahan and Weinberg proposed the now well-known "hallmarks of cancer" in 2000 [11]. This view presented a framework governing the transformation from a normal to disease state, which occurs by successive acquisition of hallmark properties. Rather than each cancer being considered a separate disease, the hallmarks aim to delineate the differences between healthy and cancerous cells. This cancer systems hypothesis has since been expanded to encompass advances in research [12], effectively further displaying the complexity of the disease.

In many cases the functionality of genetic aberrations remains unknown. Rather than attempting to map each of these genetic anomalies to a specific deregulated oncogene or tumor suppressor gene, the shared hallmark characteristics suggest a convergence of these many varied mutational profiles into a less complex phenotypic space, governed by fewer underlying principles [13]. Given that proteins are the mediators of cellular response to genetic cues, and thus one step closer to the observed phenotype, the complexity at the protein level might also be expected to be less than that at the genetic level [14].

Practically speaking, approaching these characteristics is typically done by studying the pathways associated with particular cellular processes. A discrete number of pathways can serve as an organizing principle with respect to the many players in the proteome. Indeed, thousands of mutations can be detected in a single tumor sample [15] and can be mapped to a far smaller number of processes, of the order of tens [16]. This linear paradigm appeals due to its simplicity but also in particular since pathways are by definition composed of many components, providing multiple targets for drug development rather than needing to target a single oncogene or tumor-suppressor related to a genetic defect. However, due to the complexity of interactions among the different components (e.g. cross-talk, feedback mechanisms), the consequences of targeting various nodes are not equivalent (Fig. 1.1, taken from [17]).

In order to understand how cancerous cells decipher and process internal and external signals as well as to elucidate the mechanisms behind phenomena such as resistance or metastasis, it is critical to understand how to interpret the various omics datasets together in order to gain a deeper understanding of these complex processes. This begins perhaps with the profiling of patient tissue samples and cell lines using deep sequencing and proteomics to understand the relationships between gene-level perturbations and protein-level phenotype.

1.1.2 Large-scale studies in cancer research

As mentioned earlier, the genomics field has seen the development of high-throughput next generation sequencing technologies such as whole exome sequencing (WES), whole genome sequencing (WGS), RNA-seq and ChIP-seq arrays over the past two decades which have been critical in providing genome-wide data and thus in shifting toward systematic analyses. Although genome-wide association studies (GWAS) - the detection of SNPs associated with a phenotypic trait - have identified numerous relevant genes as being

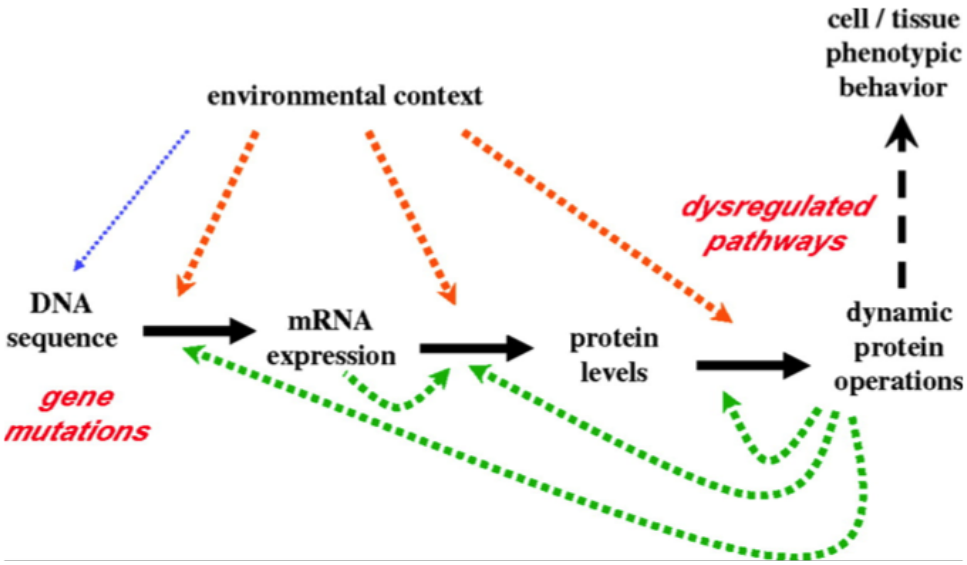


FIGURE 1.1: Depiction of the effect of environmental context on the various levels of biological regulation and how these interact toward a cell or tissue phenotype. Environmental context can intervene at the genetic, transcriptional, translational and post-translational levels. The protein level (both in terms of concentration as well as post-translational modifications can be considered to convey the effective result of the cumulative regulatory levels and is the most proximal to phenotypic behavior, thus justifying its quantification as a readout of cellular processes.

linked to investigated traits, they still fail to explain much of the diversity in genetic traits, such as height as well as in common complex diseases such as cardiovascular disease or diabetes.

In an attempt to make sense of the immense amount of sequencing data, the passenger-driver paradigm emerged as providing a way to classify cancer-related mutations. Driver mutations are those conferring a growth advantage to a cell and thus critical to malignancy. Passenger mutations represent those that occur as a result of the genomic instability in cancer cells. These passenger mutations were considered not to augment the fitness advantage of a cell and thus have no effect [18]. Within this paradigm, GWA studies were expected to uncover associations between SNPs and the cancerous phenotype, where a small number of driver mutations stand out as the main culprit. The identified genes linked to cancer would in turn determine novel treatment avenues. This approach has been successful in certain cases. For instance, the identification of a mutation in complement factor H in a study focused on macular degeneration linked inflammation to this disease, and targeting this pathway has shown therapeutic success [19]. However, for the most part it has become apparent that cancer is driven not by a single mutation but is likely the result of a larger number of mutations. Furthermore, even the same mutation may have different phenotypic outcomes depending on the state of the network the expressed protein resides in [20]. Indeed, there are several limitations concerning GWA studies, one of which is the difficulty in detecting low frequency mutations, which comprise the majority of the cancer landscape [21].

With this in mind and taking into account once again the broad variety of cellular processes that cancer impacts, several large-scale consortia such as The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) have led efforts to sequence and characterize the genomes of cancerous cells and tissues using a multi-platform approach: whole genome sequencing, miRNA expression, RNA-seq, and DNA copy number variation [22], [23]. TCGA thus provided the first large scale integrative view of the genetic landscape of ovarian cancer, identifying mutated genes and suggesting differing therapeutic strategies depending on disease subtype, or "subtype-stratified care".

The many genomic and transcriptomic efforts to sequence cancer systems have shown that mutations leading to loss-of-function are more prevalent than those related to gain-of-function [24], the latter being the ones for which a drug may be developed (e.g. kinase inhibitors such as imatinib, erlotinib, etc.). Therefore, the number of therapeutic directions that can be explored based on genetic mutational profiles is limited. Moreover, the majority of mutations are found in non-coding regions, the functional effects of which remain poorly understood. Both of these findings support the need for integrating information at both the genetic and proteomic level in order to identify therapeutic targets.

Genomic data is typically projected onto known interaction maps and functional conclusions are made based on this information. While this background data might be true in the context in which it was generated, it does not necessarily reflect the context of another cancer cell, let alone a tumor. It is therefore also critical to stress the relevance of matched genetic and phenotypic data, which is where modern proteomics takes on an important role. Computational platforms such as ReKINect [25] bring together genomics and signaling networks in an effort to identify mutations causing rewiring of network dynamics and thus providing functional information associated with these mutations.

1.2 Signaling networks as drug targets ?

While the importance of a comprehensive catalog of the genetic basis of cancers should not be dismissed, it is questionable how much it will contribute in terms of therapeutic advance. The outcome of these datasets is the suggestion of therapeutic targets based on deregulated genes which for the majority are already known [14]. However, the targets of drugs are typically proteins and it is signal transduction proteins that are behind the cancer phenotype we observe and the associated malignant behavior. Methods for the treatment of cancer (excluding surgery) can be classified grossly into the following categories: chemotherapy, radiotherapy, targeting oncogenes, targeting processes harnessed by cancer cells (e.g. angiogenesis), or immunotherapy (most recently, CAR T-cell therapy). Chemotherapy and radiotherapy both target cells with increased replication, but cause severe adverse effects as they are non-specific, targeting even healthy cells that are dividing. This is not the case for drugs aimed at interfering with oncogenes. Indeed, kinase-specific drugs such as imatinib (specific to the tyrosine kinase domain of abl) or trastuzumab (targeting ErbB2 in breast cancer cells) have been successful in the clinic, even though in the long-term many patients face relapse due to

resistance to the inhbitors. If the aim based on genetics is to design novel drugs based on targets rather than on phenotypic outcome, then there is an additional biological layer separating these data and drug targets, the effects of which need to be taken into account. Indeed, as mentioned above, the correlation of mRNA levels and protein is far from perfect (Pearson correlation coefficients in the literature range from approximately 0.5 to 0.7) thus suggesting a supplementary layer of regulation that is not yet fully understood.

The work described in Chapter 4 attempts to link the migratory phenotype of triple negative breast cancer cells to signaling occurring as a response to a wounding, a "signaling perturbation" rather than a genetic perturbation, as is the case in Chapter 3. The signaling we identify may provide additional avenues for the development of cancer therapeutics and will be discussed in more detail in Chapter 4. During this work the lack of known functionally annotated phosphorylation sites was striking and suggests that there is much to be discovered regarding the functional roles of phosphorylated sites.

1.3 Chromosomal rearrangements in cancer

With the advent of the genomic revolution the landscapes of genomic alterations became increasing well-characterized, and the mass of data generated due to the large number of identified aberrations uncovered meant that system-wide questions could be addressed (see GWAS studies in section 1.1.2). Examples of point mutations associated with malignancy are BRAF V600E, as well as mutations in p53 and in src. Next-generation sequencing allows the identification of genomic modifications such as single-nucleotide polymorphisms, small insertions and deletions. The more recent development of RNA-seq has exposed the frequency of chromosomal aberrations across many tissue and cancer types [26], and their potential as oncogenes and possible drug targets has emerged. The following subsections provide an introduction to chromosomal rearrangements, their oncogenic potential and their impact on protein signaling networks, and thus why more attention should be given to their role in cancer.

1.3.1 The Philadelphia chromosome

In the early 20th century Theodor Boveri proposed that chromosomal abnormalities seen in tumor tissues were the cause of transformation toward malignant entities, thereby linking genes with cancer [27]–[29]. Only in 1959 did the Philadelphia case (the fusion of BCR and ABL1) become the first chromosomal aberration to be consistently associated with a human disease: the hallmark of chronic myelogenous leukemia (CML) is now known to be the t(9;22)(q34;q11) chromosomal translocation [30]. It is now indeed widely known that chromosomal aberrations and impaired DNA damage repair apparatus are critical in tumor initiation and progression, thereby linking genetics to phenotype [31], [32]. The BCR-ABL rearrangement provides an ideal example of a driver for cancer based on a genomic alteration, which has spurred the development of the first targeted kinase inhibitor for the treatment of CML.

The BCR-ABL1 gene fusion (the result of which is also known as the Philadelphia chromosome) has been identified in three different forms, associated with different forms of leukemia depending on the precise location

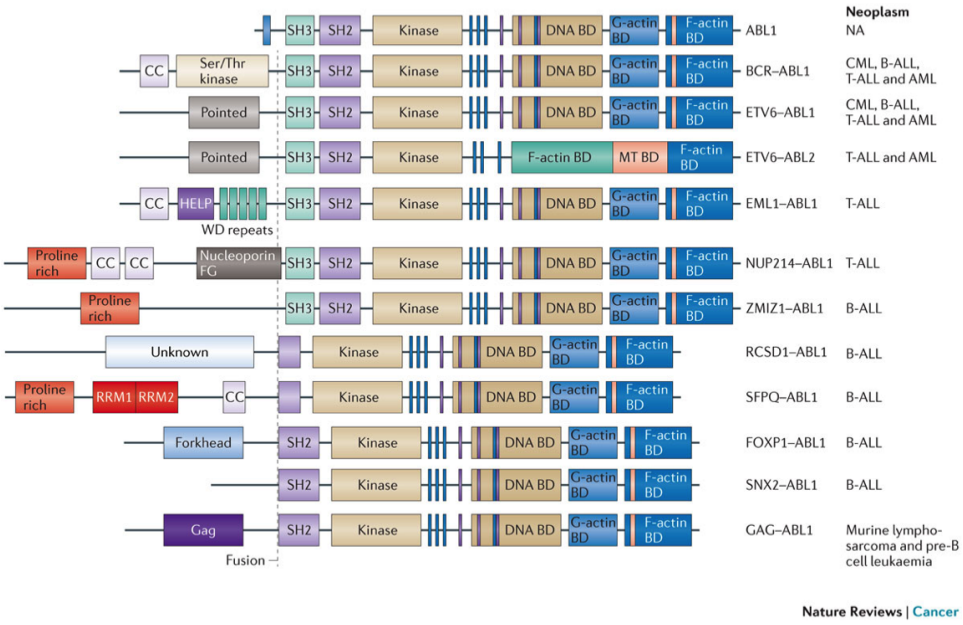


FIGURE 1.2: Image taken from Greuber *et al.*. Illustration of the diverse ABL1 and ABL2 fusion proteins resulting from chromosome translocations with various partner genes, and leading to various pathological phenotypes.

of the breakpoint : P210 BCR-ABL1 as a marker of CML [33]; P190 BCR-ABL1 (also known as p185) in many cases of adult and childhood B cell acute lymphocytic leukemia (ALL and B-ALL, respectively) [34], [35]; and P230 BCR-ABL1 in neutrophilic CML and CML [36], [37]. The exact location of the breakpoint results in different response to chemotherapy, e.g. patients harboring the p210 isoform respond well to treatment whereas others develop resistance to tyrosine kinase inhibitors. Several kinase inhibitors have been developed targeting the tyrosine kinase function of ABL1, notably imatinib (Gleevec; Novartis), dasatinib (Sprycel; Bristol-Myers Squibb) and nilotinib (Tasinga; Novartis). Imatinib has been amongst the more successful treatments of chronic phase CML and exerts its effect as an ATP competitor, but up to a third of patients develop resistance due to point mutations in the kinase domain of the BCR-ABL1 complex [38]. Moreover, the drug is less effective for patients presenting with alternative BCR-ABL fusions. Strategies to combat resistance to these drugs has been to target key proteins downstream of BCR-ABL [39]. Moreover, ABL has been identified as a fusion partner with a number of other genes, and associated with different clinical outcomes (Fig. 1.2).

1.3.2 A focus on the role of tyrosine kinases as translocation partners

Many of the well-characterized fusion genes have been identified as involved in leukemic malignancies. This could be due to the amenability of leukemic samples and cell lines, as well as a lack of systematic cytogenetic analysis. Fusions have also been identified in solid tumors, albeit at a lower rate. In particular, a non-negligible number of human kinase genes in leukemic as well as solid tumors have been found to be rearranged and involved in cancer

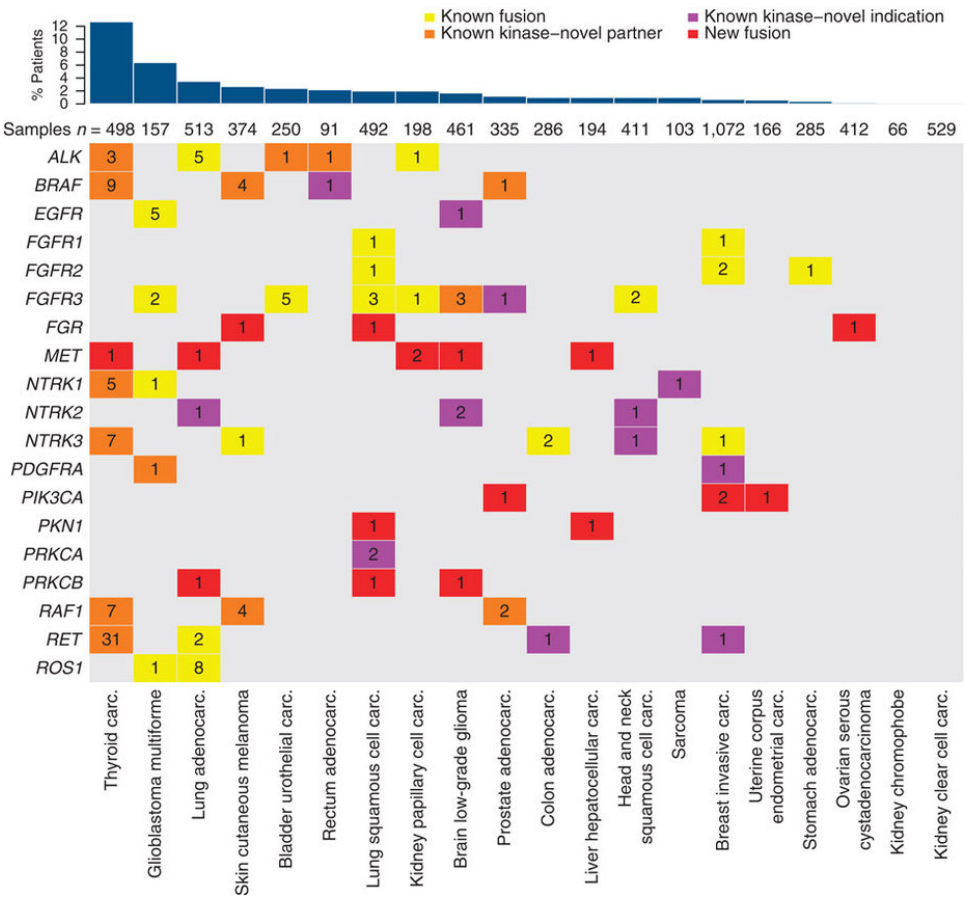


FIGURE 1.3: Image taken from Stransky *et al.*. Depiction of the recurrent kinase fusions arising in various solid tumor types. Thyroid carcinoma (leftmost) is the tissue type for which the highest percentage of (TCGA) patient samples contain a kinase rearrangement (bar plot). The color coding indicates the type of fusion depending on its recurrence as well as previous knowledge of the kinase fusion partner.

(Fig. 1.3) [26], [40], [41]. Moreover, the drug design strategy that led to the development of successes such as imatinib and dasatinib have inspired small molecule compounds with promising results for solid-tumor cancers, e.g. crizotinib for ALK raerrangements [42], [43].

Trends in the association of two partner genes and the types of aberrant behavior they result in have been found when comparing several kinase fusions [41]. In most fusion proteins with a functional kinase domain, the latter is located C-terminally and inhibitory elements have a tendency to be found at the N-terminus. In order to retain kinase activity, the intact C-terminal kinase domain is conserved and the partner gene nearly always supplants the N-terminal portion of the kinase protein. In the case of fusions where one partner is a receptor tyrosine kinase (RTK), the extracellular ligand-binding domain originally belonging to the RTK is thus often lost in the final fusion product. This is also the case for the inhibitory domains found in most tyrosine kinases (TK), which stabilize the inactive kinase conformation and thereby dampen activity when stimuli is absent. This pattern generally results in the expression of the fusion protein being driven by the promoter of

the 5' partner gene, the latter is also often critical for controlling oligomerization and thus activation of the kinase domains. This may lead to cases where the kinase domain becomes constitutively active in a cell positive for the translocation, whereas the same cell type without the rearrangement may display low or no expression of the wild-type kinase under normal conditions [44]. For instance, upon induction of BCR-ABL expression in mice CML occurred only if the expression was induced specifically in haematopoietic stem cells. A BCR-ABL fusion gene under the control of a different promoter generated other types of neoplasms [40], [45], [46].

There are several ways in which the kinase moiety can be deregulated in the context of chromosomal rearrangements. Activation of tyrosine kinase domains within fusion products can occur if the partner gene expresses multimerization domains, thereby bringing kinase domains into proximity and mimicking RTK activation, e.g. the PNT domain of ETV6, which associates with ABL, PDGFRB, JAK2, and TRKC [47]–[50]. Coiled-coil domains appear to be most frequently used to drive oligomerization in TK fusions (60% of tyrosine kinases harbor a coiled-coil domain, vs. 9% in the entire human proteome) [41], however a number of oligomerization domains have been detected in unique fusions. Furthermore, the degree of oligomerization can vary from dimers to polymeric structures [51]–[53] and it has been suggested that not all forms possess transformative potential [49]. Deletion of the BCR-ABL oligomerization domain was shown to severely limit the fusion product's tyrosine kinase activity and in vitro transformative capacity [54].

The partner gene can provide additional advantageous functionalities, such as targeting the fusion product to the nuclear pore complex, recruiting additional proteins resulting in a larger complex [55], recruiting chaperones mediating protein stability [56], and the recruitment of the wild-type form of the partner protein thereby exacerbating the effect of the translocation. As a consequence of these changes in the regulation of the kinase domains, recurrent cancer-specific rearrangements drive the cell toward a diseased state and provide it with a growth and proliferation advantage [57], [58] ultimately leading to transformation, marking the transition from normal to malignant cells.

1.3.3 Transforming capacity of fusion kinases

In 1988 Daley & Baltimore [59] showed that p210 bcr/abl was capable of transforming hematopoietic cell types [46]. Similarly, in 2005 a study by Warner*et al.* [60] demonstrated full leukemic conversion of normal human hematopoietic cells by introduction of TLS-ERG using retroviral transduction techniques. This led to the spontaneous occurrence of further genetic and epigenetic alterations, providing evidence for multiple cooperating events in human leukemogenesis. Indeed, although the direct effects of TLS-ERG expression alone were insufficient to induce complete transformation, extended culture of these cells resulted in the acquisition of additional changes and gave rise to a fully transformed hematopoietic cell line from primary cells. Interestingly, a requirement for maintenance of telomere activity emerged, and was shown to be a critical second event following introduction of TLS-ERG in order for cells to achieve full transformation (see Fig. 1.4) [60].

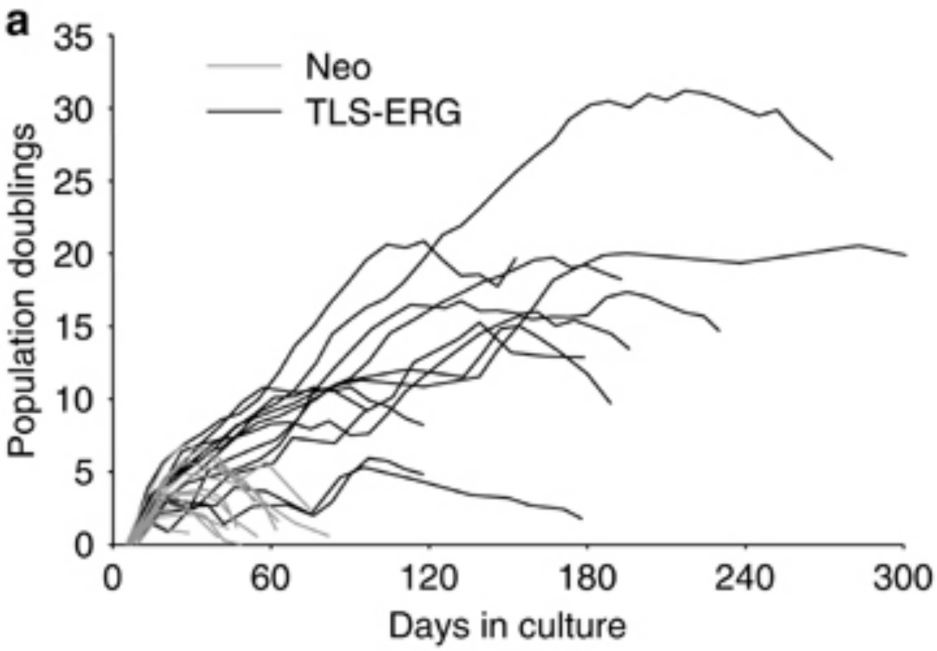


FIGURE 1.4: Adapted from Warner *et al.*. TLS-ERG fusion expressed in human cord blood cells, compared to "Neo" negative control (no fusion). After a certain amount of time in culture, the fusion-containing cells show increased proliferation.

With this in mind it is interesting to note that Lundberg *et al.* (2002) presented a new model facilitating the study of transforming events by immortalization of a primary cell line by the successive introduction of SV40 early region and the telomerase catalytic subunit hTERT. Indeed, these cells show prolonged proliferative potential however are not capable of forming colonies in assays testing for anchorage independant growth [61] (Fig. 1.4).

In the case of BCR-ABL, the fusion has been shown to activate Ras via different mechanisms. In one scenario, the fusion kinase can be coupled to the MAPK pathway, capable of transformation by activating Ras [62]. The latter is frequently found mutated in tumors and disturbs the regulation of core cellular processes such as proliferation and differentiation. In any case, mitogenesis is induced by providing the cell with aberrant stimulation by the TK, which replaces signaling that would normally be tightly regulated by cytokine receptors [63]. Another example of cytokine signaling hijacking are the STAT family proteins which are involved in the regulation of cell cycle progression, where amongst others STAT5 has been shown to be activated following fusion tyrosine kinase (FTK) expression [64]. However, this seems to depend on the exact translocation, since STAT5 is not essential for transformation by BCR-ABL [65].

The replicative stress and genomic instability induced by the FK might allow for the accumulation of secondary mutations necessary for disease progression. However as fusion genes can be detected in the DNA of healthy patients, expression of the fusion itself may be insufficient for transformation and require additional mutations [66]. Another interesting point is whether these secondary mutations are a direct consequence of fusion protein activity

or arise due to the transformed state of the cell. Despite the evidence pointing toward the first situation [67]–[70], the number of genetic abnormalities in CML cells are rather low implying that the BCR-ABL fusion protein itself might not possess the anticipated mutagenic potential [44].

In normal cells apoptosis is triggered in response to DNA damage, providing a barrier to proliferation of cancerous cells. FKs allow cells to circumvent this limitation by sending survival signals to block apoptosis [71]. BCR-ABL was even shown to reverse the apoptotic process induced by cytokine deprivation, thereby enabling transformation [72].

Fully transformed cells show a disruption of DNA repair mechanisms, thereby maintaining genomic instability which is thought to be involved in the further progression of the disease to an advanced state [73]. Moreover, the fusion kinase is thought to cause 'oncogene addiction' due to the suppression of the apoptotic response ensuring further growth despite increased DNA damage due to genomic instability. Oncogene addiction is the dependance of a cancer cell on an oncogenic for its malignant properties, and targeting this oncogene is a rationale for targeted therapy [74]. The PI3K pathway is particularly interesting in this context as it is involved in both survival and mitogenic signaling [75]. Some or all three of the MAP kinase pathways (ERK, JNK, and p38) have been noted to be activated in fusion kinase-transformed cells [76]–[78]. In contradiction, these pathways have also been reported not to be activated for different FTKs or to have opposite effects in different cell types [79]–[81].

In conclusion, the cellular states critical for transformation as well as the aftermath of transformation have been given much attention, however the transformation event itself remains elusive. Induction of mitogenesis, replicative stress, undermining of DNA repair and apoptosis are likely to be the conditions critical for transformation, but the causality between them is currently still unclear and, given the encouraging results common to several FTKs, it is tempting to believe that a pattern will emerge from studying as broad a panel of fusion kinases as possible. However, in light of the above contradictions, careful selection of the appropriate model system (cell lines, organ, tissue) will be critical to make biologically relevant conclusions and advance our understanding of progression toward a cancerous cell state.

1.3.4 Therapeutic potential

The earliest cancer therapeutics were targeted at the main deregulated phenotypic aspects of cancer cells, but are also highly toxic for the patient. For instance, the Achilles heel of cancerous cells proliferating in an uncontrolled manner is at the moment of their replication. Since these cells replicate at a higher frequency than healthy cells, targeting drugs at the replication machinery has been a strategy to hit malignant cells. Healthy cells still replicate albeit at a lower rate, and thus chemotherapies based on this mechanism show high toxicity and carry many adverse effects for patients.

The advent of high-throughput sequencing brought about the concept of personalized medicine, where patients are administered a drug regimen tailored to their personal mutational profile. In the case of BCR-ABL in blood-related cancers, a physician might prescribe a different duration of imatinib treatment before switching to an alternative kinase inhibitor depending on the exact genetic fusion breakpoint. This shift toward the development of

personalized therapeutics has been heralded as the next breakthrough in cancer therapy, however its success has been limited. Indeed, cancer cells tend to develop resistance to drugs by e.g. mutation of a kinase's active site resulting in decreased inhibitor affinity. Network medicine was first described as a systems approach to understanding disease taking advantage of network properties [82]. Going one step further, it has been speculated that a multi-pronged therapeutic approach where several network nodes are targeted either simultaneously or in succession could, if adequately dosed, address the limitations of current personalized medicine attempts [83].

As discussed in more detail in Chapter 2, an efficient gene editing tool such as CRISPR-Cas9 could provide potentially enormous therapeutic advances. Still, this technology has several limitations in common with siRNA-mediated gene silencing (RNAi), such as difficulty in delivery and off-target effects. Additionally, the permanent nature of these genetic modifications render these issues perhaps even more critical than in the case of RNAi. Imagining a context where all these limitations were insignificant, this technology would render genetic diseases amenable to permanent cure. This would be the only way to directly target the many chromosomal aberrations, such as SNPs or rearrangements that are currently being identified in high-throughput patient tumor profiling initiatives.

1.4 Fusion proteins and signaling networks

Transformation in the presence of kinase fusion genes is in reality the consequence of the de-regulation of signalling pathways that control cell proliferation and inhibition of apoptosis. TK fusions, via their kinase moiety, are thought to directly or indirectly elicit events leading to transformation such as genomic instability, deregulated cell growth, as well as subversion of the apoptosis program for cells with DNA-damage [84]. As a result of the history of discoveries involving chromosomal translocations, a large portion of the work focuses on the understanding of BCR-ABL, as well as on a few alternative fusions found to direct cells toward a leukemic state. Two recent proteomic studies focused on the signaling differences between isoforms of the BCR-ABL fusion [85], [86]. The two studies found changes in the signaling network, despite the kinase domain being preserved in both isoforms.

Despite the confounding literature and the lack of large-scale studies, some commonalities emerge among tyrosine kinase fusions implicated in leukemia [nicely reviewed in [44]]. The preservation of the oligomerization domain allows autophosphorylation of the protein, through which downstream signal effectors are activated which in turn lead to transformation in hematopoietic cells. A necessary condition for transformation is the subversion of normal constraints on cell cycle control, allowing for unrestricted entry into mitosis. Studies show that induction of mitogenic pathways due to the fusion protein are central to transformation. Replicative stress then creates a state of genomic instability where accumulation of mutations and further DNA-damage can occur. In turn, this then provides the context for a modification in the equilibrium between pro- and contra-oncogenic states. Finally, survival signals are triggered by the fusion kinase to avoid entry into apoptosis. Some pathways therefore appear to be commonly de-regulated: the phosphatidylinositol-3-kinase (PI3K) pathway, mitogen-activated protein

kinase (MAPK) pathways, STAT-family proteins, and NF-κB. Some examples of the conflicting studies regarding the process of fusion kinase-mediated transformation follow.

1.4.1 Proteomic studies of chromosomal translocations

The majority of studies to date consider the fusion protein and its effects downstream of a pathway of interest. Recently, there has been an increase in experiments using large-scale proteomics and phospho-proteomics to gain a more quantitative understanding of the effect of translocations although the bulk pertain to BCR-ABL1 and the effect of associated tyrosine kinase inhibitors [87]–[92]. The use of mass spectrometry in addition to other high-throughput data sources provided the ability to accurately quantify the effects of kinase inhibitors or combinations thereof on the state of cells, thus suggesting a mechanism for drug interactions and activity. Fairly few are harnessing the power of this fast-developing technology to obtain a global view of the signaling occurring due to the expression of the many less well-known fusion kinases, the path to cellular transformation, and the maintenance and progression of cancerous cells [93]–[96]. As discussed in a review by Turner [44], there are likely commonalities among these proteins that remain to be discovered.

While cases of patients harboring novel chromosomal translocations are constantly discovered, it will be critical to delve into a deeper understanding of the mechanisms governing their creation, development, and effects in order to translate this wealth of data into therapeutic strategies.

Two recent studies aimed at understanding the difference in expression of the BCR-ABL p210 and p190 isoforms with respect mainly to signaling, relevant due to the difference in clinical outcomes [85], [86]. While a number of studies had used mass spectrometry to study the BCR-ABL rearrangement these were the first to perform a quantitative systematic analysis in the same cellular background. The studies both concluded that although the isoforms have similar kinase activation, there are a number of differentially regulated interactors and downstream signaling partners, namely members of the Stat as well as the Src families.

In order to study the effects of proteins on the cell, the cellular context as well as the method of introducing the perturbation is key to maintaining biological relevance of the study. Rather than overexpressing fusion constructs, the low frequency of translocations is preserved when using a nuclease to generate specific double strand breaks [97]. This has been done with engineered Zinc-finger nucleases as well as TALENs [98] [99], and more recently using the CRISPR-Cas9 system [100]–[102]. Choi *et al.* created several rearrangements (CD74-ROS1, EML4-ALK, KIF5B-RET) as did Torres *et al.* (EWSR1-FLI1 and RUNX1-ETO). The fusions were induced in various cell lines at a low frequency ($<$10% of targeted cells carried the rearrangement), and were shown to be at precisely the desired intronic location, demonstrating the success of using nucleases to generate double strand breaks and subsequent isolation of the desired translocations. Moreover, Maddalo *et al.* demonstrated that this strategy is also applicable *in vivo*. The low frequency of rearrangement is certainly a limitation to the experimental study of fusions using this method, but also likely reflects the physiological conditions in which fusions occur. Moreover, the increase in papers using CRISPR-Cas9 in genome-wide

screens [103] confirms the use of this method for large-scale studies such as mass spectrometry-based proteomics. It should be noted however that off-target effects of nucleases are recognized but not yet fully understood and are a limitation of the use of nucleases that should be taken into account (discussed in Chapter 2).

1.4.2 Mass spectrometry-based proteomics

Where next generation sequencing identifies abnormalities in the genomes of cancerous systems, high-throughput proteomics allows a quantitative assessment of their consequences one step closer to the phenotypic level. The cancer community still lacks a robust understanding of how genomic events impact the proteome and phosphoproteome of a cell, resulting in phenotypic changes that are then identified as disease. Measurement directly at the proteomic level is essential for a complete understanding of the biological systems at hand as data generated by genomics or gene expression studies does not provide an accurate representation of protein abundance nor of their dynamics.

For the large-scale analysis of proteins in 'discovery mode', liquid chromatography-tandem mass spectrometry has emerged as the preferred technology. The typical "bottom-up" proteomics workflow involves the tryptic digestion of proteins into peptides, which are then separated by reverse phase liquid chromatography. Voltage applied to the eluate ionizes peptides, and these ions are transferred to the gas phase for direct injection into the mass spectrometer. A detailed description of the electronics residing inside modern mass spectrometers and the physico-chemical reactions occurring within are complex and beyond the scope of this introduction. Briefly, the mass-to-charge (m/z) ratio of the ionized peptides (precursor ions) is measured in a mass analyzer (common mass analyzers include the Orbitrap, ion trap or time-of-flight (TOF)). In the case of the Orbitrap, ions are trapped around an electrode, and the frequency of rotation and oscillation of the ions around the electrode are recorded over time. Using a Fourier transform, these measurements are then converted from a time into a frequency domain from which the mass-over-charge ratio of an ion can be inferred. Once the m/z precursor peptides are measured (MS), a subset of ions are chosen based on abundance and charge and selected for further fragmentation (fragment ions). To fragment ions, ions of desired mass are isolated and excited into colliding with gas ions (e.g. helium) causing fragmentation of the precursor ion's chemical bonds. These fragmented ions are then sent to a mass analyzer, where their m/z and intensity are acquired (MSMS). The data acquired for these precursor ion-fragment ion pairs is then searched against a database of spectra in order to identify the amino acid sequence of the tryptic peptide. Quantification can be achieved e.g. by comparing the abundance of ions (label-free quantification) or by using isotopic labeling, whereby stable amino acid isotopes are incorporated into the proteins of a sample to compare with, and then mixed at a known ratio.

The advent of "shotgun" mass spectrometry-based proteomics has provided researchers with the capacity to measure thousands of proteins within a single sample, a drastic increase from the well-established antibody-based proteomic techniques such as reversed phase protein arrays where antibody availability limits the scope of an experiment. Additionally, antibody-based

quantitation methods are typically based on a single epitope, yielding only a semi-quantitative picture of protein levels. Therefore, a focus of developing mass spectrometry-based methods has been precise and accurate protein quantitation, some well-established approaches being isotope labeling with amino acids in cell culture (SILAC), isobaric labeling (iTRAQ, TMT) and label-free quantification (LFQ) [104]–[107].

Signaling via post-translational modifications (PTM) such as phosphorylation, glycosylation and ubiquitination, as well as their cross-talk, are critical in modulating the cell's response to external and internal stimuli. Importantly, analysis of phosphorylation is critical in understanding the rewiring of protein networks upon genetic alterations. PTM enrichment strategies allow the measurement of various sub-proteomes such as the phospho-proteome or glycoproteome providing an additional dimension to global proteomics. Generating quantitative data for PTM sub-proteomes is an important addition to the omics list, since this post-translational information can by definition not be gleaned from the genetic or transcriptomic levels. More importantly, changes in levels of post-translational modifications reflect properties of a protein, such as a change in protein activity or conformation.

Phosphorylation in particular is a well-studied PTM and the phosphorylation state of proteins in a cell or system is measured via mass spectrometry of the phosphorylation-enriched proteome, informing on the activity, localization and stability of the phosphorylated protein. This is typically done before and after a biological perturbation (drug, stimulation, etc.) in order to understand the changes occurring in the cellular network of modified and unmodified proteins. This can to some extent also inform on the upstream kinase activity, although it should be underlined that this method provides a measure of kinase activity via a proxy and is therefore an indirect measurement. Other processes such as additional PTM regulation (e.g. ubiquitination, glycosylation), may lead to uncorrelated kinase and phosphorylation-site measurements, complicating the analyses. The combination of multiple types of mass spectrometry data can fill the voids left by individual experimental setups: for example, measurement of the entire proteome, the phospho-enriched proteome, and immuno-enrichment of phospho-tyrosine residues have been combined [108], [109].

Proteomics has demonstrated its power in identifying and quantifying proteins and biomarkers in cells as well as in different tissue types. With the advent of ever more sensitive and fast instruments the focus has shifted toward a more systems approach profiling proteins across many perturbation states and time points. While this has certainly furthered our understanding of the roles of proteins involved in e.g. disease vs. healthy states, there is much room for improvement. For instance, the quantification methods touched upon above are for the most providing relative protein quantification. Robust absolute quantification will likely be an important advance for the proteomics community in the future.

1.5 Proteomic data integration strategies

Due to the biases each technology and experimental design is associated with, the collection and analysis of complementary datasets is one approach in an

effort to move toward a more complete representation of the studied system of interest. With this in mind, the Clinical Proteomic Tumor Analysis Consortium (CPTAC) [24] was created in order to add an additional omics dimension to the large-scale genomic cancer datasets, and thereby build on previous efforts. The CPTAC uses multiple stages of mass spectrometry-based proteomics: a discovery phase where global proteomic profiling is performed on a select subset of tumors that have been genetically characterized by TCGA to identify proteins of interest, and a second validation phase during which targeted assays are performed in order to monitor the proteins of interest, e.g. potential biomarkers or drug targets. The first phase integrates genomic prior data into proteomic analysis by providing an improved annotation of the reference genome/proteome, and by providing a list of genetic mutants on which to confirm peptide-level mutants.

1.5.1 Data heterogeneity

Integrating data from multiple sources presents a major difficulty due to the inherent differences of the individual technologies. In some cases data might be available for a population of cells (e.g. RNA-seq, mass spectrometry), whereas others provide an output at a single-cell level (e.g. imaging, single-cell mass cytometry). Another difference is sensitivity; discovery proteomics focuses more on the broadest identification of the system's components, in order to lay a network map of proteins in the system. Therefore, a lacking protein identification does not necessarily signify that the entity is truly lacking. Rather, its abundance was not sufficient in order to be confidently identified. Targeted mass spectrometry techniques on the other hand are capable of high sensitivity and may be used to determine the presence or absence of a peptide in a given sample.

Using multiple techniques side-by-side separately to provide orthogonal evidence supporting a particular finding is nowadays expected to support major findings, however integrating large datasets in a systematic manner is far from straightforward. Additionally, due to the features of the various data types, different modeling approaches are required.

1.5.2 Proteogenomics

One example of successful integration is that of sequencing and proteomics at a qualitative level, or proteogenomics [110], [111]. In this case, exome sequencing or RNA-seq data is used to create a sample-specific list of gene and therefore protein sequences, which can then be used to query proteomic data by feeding this novel information into a reference database. This is especially well-suited for patient tissue analysis where a time-course setup is difficult to implement and it is therefore difficult to study the evolution of the disease. In order to gain as much insight as possible when analyzing patient samples, proteomic and genomic measurements may be acquired from the same tissue sample and the patient's genetic profile can be used for a more informed proteomic analysis. This data can then be overlaid on known protein-protein interactions to infer regulatory networks. Indeed, a genetic alteration (such as SNPs, frameshifts and splice isoforms) leading to an alternative protein isoform would be missed in a proteomic experiment without inclusion of this variant in the search database. Mertins *et al.* have found that detection

of genomic variants in mass spectrometric data is sparse, partly due to the incomplete coverage of proteins by the individually measured peptides [110]. Rather than use mass spectrometry to confirm genetic alterations directly as alternative protein isoforms, genetic variants may be identified and subsequent changes in protein expression or PTM regulation can be determined. It is in this context of increased knowledge of the chemical space of protein isoforms that we apply this technique in the projects described in part II.

1.6 Context of this thesis

The studies in part II are examples of creating a cellular perturbation akin to those occurring in a disease setting and subsequently sampling this altered network. Chapter two is an introduction to chapter three in which the effect of a genetic perturbation, the generation of a chromosomal translocation, is assessed at the protein abundance, signaling and phenotypic levels. Chapter four perturbs a cell population directly at the level of signaling using a 2D wound assay and identifies signaling changes associated with increased migration. Although different in many aspects, the two studies share the underlying assumption that a perturbation is not an isolated event leading to a single outcome. Rather, the effects reverberate throughout the network of interacting proteins, the consequences of which may be the evolution towards an increasingly malignant state.

1.7 References

[1] J. Monge, M. Kricun, J. Radovčić, D. Radovčić, A. Mann, and D. W. Frayer, "Fibrous Dysplasia in a 120,000+ Year Old Neandertal from Krapina, Croatia.", *PLoS ONE*, no. 6, e64539, 2013.

[2] World Health Organization, "World Cancer Report 2014.", World Health Organization, Tech. Rep., 2014.

[3] T. Beadle, "Genetic Control of Biochemical Reactions in Neurospora.", *Proc Natl Acad Sci U S A.*, vol. 27, pp. 499–506, 1941.

[4] U. Sauer, M. Heinemann, and N. Zamboni, "Getting Closer to the Whole Picture.", *Science*, vol. 316, no. 5824, pp. 550–1, 2007.

[5] N. Kumar, R. Afeyan, H. Kim, and D. Lauffenburger, "Multipathway model enables prediction of kinase inhibitor cross-talk effects on migration of Her2-overexpressing mammary epithelial cells.", *Molecular Pharmacology*, vol. 73, no. 6, pp. 1668–78, 2008.

[6] H. Chuang, E. Lee, Y. Liu, D. Lee, and T. Ideker, "Network-based classification of breast cancer metastasis", *Molecular Systems Biology*, no. 140, 2007.

[7] Lander E, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, and Funke R, "Initial sequencing and analysis of the human genome International Human Genome Sequencing Consortium", *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[8] J. C. Venter, M. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, and Gocayne, "The Sequence of the Human Genome.", no. 5507, pp. 1304–51, 2001.

[9] International Human Genome Sequencing Consortium, "Finishing the euchromatic sequence of the human genome.", *Nature*, no. 7011, pp. 931–45, 2004.

[10] M. A. Dawson and T. Kouzarides, "Cancer epigenetics: From mechanism to therapy.", *Cell*, vol. 150, no. 1, pp. 12–27, 2012.

[11] D. Hanahan and R. A. Weinberg, "The Hallmarks of Cancer", *Cell*, vol. 100, pp. 57–70, 2000.

[12] ——, "Hallmarks of Cancer: The Next Generation.", *Cell*, no. 5, pp. 646–74, 2011.

[13] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler, "Cancer Genome Landscapes.", *Science*, vol. 339, no. 6127, pp. 1546–58, 2013.

[14] M. B. Yaffe, "The Scientific Drunk and the lamppost: Massive Sequencing efforts in cancer Discovery and Treatment.", *Science Signaling*, vol. 6, no. 269, pe13, 2013.

[15] L. Alexandrov, S. Nik-Zainal, D. Wedge, J. Aparicio, S. Behjati, and A. e. a. Banking, "Signatures of mutational processes in human cancer.", *Nature*, no. 7463, pp. 415–21, 2013.

[16] L. D. Wood, D. W. Parsons, S. Jones, J. Lin, T. Sjöblom, R. J. Leary, D. Shen, S. M. Boca, T. Barber, J. Ptak, N. Silliman, S. Szabo, Z. Dezso, V. Ustyanksky, T. Nikolskaya, Y. Nikolsky, R. Karchin, P. A. Wilson, J. S. Kaminker, Z. Zhang, R. Croshaw, J. Willis, D. Dawson, M. Shipitsin, J. K. V. Willson, S. Sukumar, D. G. Ballinger, A. B. Sparks, J. Hartigan, D. R. Smith, E. Suh, N. Papadopoulos, P. Buckhaults, S. D. Markowitz, G. Parmigiani, K. W. Kinzler, V. E. Velculescu, and B. Vogelstein, "The Genomic Landscapes of Human Breast and Colorectal Cancers.", *Science signaling*, vol. 318, no. 5853, pp. 1108–13, 2007.

[17] P. K. Kreeger and D. A. Lauffenburger, "Cancer systems biology: a network modeling perspective.", *Carcinogenesis*, vol. 31, no. 1, pp. 2–8, 2010.

[18] M. R. Stratton, P. J. Campbell, P. A. Futreal, D. Von Hansemann, and T. Boveri, "The cancer genome", *Nature*, vol. 458, no. 7239, pp. 719–24, 2009.

[19] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff, and H. Parkinson, "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.", *Nucleic Acids Research*, pp. D1001–6, 2014.

[20] P. Creixell, E. Schoof, J. Erler, and R. Linding, "Navigating cancer network attractors.", *Nature Biotechnology*, vol. 30, no. 9, pp. 842–8, 2012.

[21] D. Urbach, M. Lupien, M. R. Karagas, and J. H. Moore, "Cancer heterogeneity: origins and implications for genetic association studies.", *Trends in Genetics*, vol. 28, pp. 538–543, 2012.

[22] TCGA, "Integrated Genomic Analyses of Ovarian Carcinoma.", *Nature*, vol. 474, no. 7353, pp. 609–615, 2011.

[23] S. NikZainal, H. Davies, J. Staaf, M. Ramakrishna, D. Glodzik, X. Zou, I. Martincorena, L. B. Alexandrov, S. Martin, D. C. Wedge, P. Van Loo, Y. S. Ju, M. Smid, A. B. Brinkman, S. Morganella, M. R. Aure, O. C. Lingjærde, A. Langerød, M. Ringnér, S.-M. Ahn, S. Boyault, J. E. Brock, A. Broeks, A. Butler, C. Desmedt, L. Dirix, S. Dronov, A. Fatima, J. A. Foekens, M. Gerstung, G. K. J. Hooijer, S. J. Jang, D. R. Jones, H.-Y. Kim, T. A. King, S. Krishnamurthy, H. J. Lee, J.-Y. Lee, Y. Li, S. McLaren, A. Menzies, V. Mustonen, S. O'Meara, I. Pauporté, X. Pivot, C. A. Purdie, K. Raine, K. Ramakrishnan, F. G. Rodríguez-González, G. Romieu, A. M. Sieuwerts, P. T. Simpson, R. Shepherd, L. Stebbings, O. A. Stefansson, J. Teague, S. Tommasi, I. Treilleux, G. G. Van den Eynden, P. Vermeulen, A. Vincent-Salomon, L. Yates, C. Caldas, L. van't Veer, A. Tutt, S. Knappskog, B. K. T. Tan, J. Jonkers, Å. Borg, N. T. Ueno, C. Sotiriou, A. Viari, P. A. Futreal, P. J. Campbell, P. N. Span, S. Van Laere, S. R. Lakhani, J. E. Eyfjord, A. M. Thompson, E. Birney, H. G. Stunnenberg, M. J. van de Vijver, J. W. M. Martens, A.-L. Børresen-Dale, A. L. Richardson, G. Kong, G. Thomas, and M. R. Stratton, "Landscape of somatic mutations in 560 breast cancer whole-genome sequences.", *Nature*, vol. 534, no. 7605, pp. 47–54, 2016.

[24] M. J. Ellis, M. Gillette, S. A. Carr, A. G. Paulovich, R. D. Smith, K. K. Rodland, R. R. Townsend, C. Kinsinger, M. Mesri, H. Rodriguez, D. C. Liebler, and Clinical Proteomic Tumor Analysis Consortium (CPTAC), "Connecting genomic alterations to cancer biology with proteomics: the NCI Clinical Proteomic Tumor Analysis Consortium.", *Cancer Discovery*, vol. 3, no. 10, pp. 1108–12, 2013.

[25] P. Creixell, E. M. Schoof, C. D. Simpson, J. Longden, C. J. Miller, H. J. Lou, L. Perryman, T. R. Cox, N. Zivanovic, A. Palmeri, A. Wesolowska-Andersen, M. Helmer-Citterich, J. Ferkinghoff-Borg, H. Itamochi, B. Bodenmiller, J. T. Erler, B. E. Turk, and R. Linding, "Kinome-wide Decoding of Network-Attacking Mutations Rewiring Cancer Signaling.", *Cell*, no. 1, pp. 202–17, 2015.

[26] N. Stransky, E. Cerami, S. Schalm, J. L. Kim, and C. Lengauer, "The landscape of kinase fusions in cancer.", *Nature Communications*, no. 4846, 2014.

[27] D. Hansemann, "Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung.", *Archiv fur Pathologische Anatomie und Physiologie und fur Klinische Medicin*, 1890.

[28] T. Boveri, "Uber mehrpolige Mitosen als Mittel zur Analyse des Zellkerns", *Verhandlungen der physicalisch-medizinischen Gesselschaft zu Würzburg*, vol. Neu Folge 35, pp. 67–90, 1902.

[29] D. W. Cleveland and A. J. Holland, "Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis.", *Nature*, vol. 10, no. 7, pp. 478–87, 2009.

[30] P. Nowell, "Discovery of the Philadelphia chromosome: a personal perspective.", *The Journal of Clinical Investigation*, vol. 117, no. 8, pp. 2033–5, 2007.

[31] A. E. Teschendorff and C. Caldas, "The breast cancer somatic 'mutaome': tackling the complexity.", *Breast cancer research : BCR*, vol. 11, no. 2, p. 301, 2009.

[32] I. B. Tan, I. Ng, W. M. Tai, and P. Tan, "Understanding the genetic basis of gastric cancer: recent advances.", *Expert Review of Gastroenterology & Hepatology*, vol. 6, no. 3, pp. 335–341, Jun. 2012.

[33] K. De Keersmaecker, J. L. Rocnik, R. Bernad, B. H. Lee, D. Leeman, O. Gielen, H. Verachtert, C. Folens, S. Munck, P. Marynen, M. Fornerod, D. G. Gilliland, and J. Cools, "Kinase Activation and Transformation by NUP214-ABL1 Is Dependent on the Context of the Nuclear Pore.", *Molecular Cell*, no. 1, pp. 134–42, 2008.

[34] H. M. Kantarjian, M. Talpaz, K. Dhingra, E. Estey, M. J. Keating, S. Ku, J. Trujillo, Y. Huh, S. Stass, and R. Kurzrock, "Significance of the P210 Versus P190 Molecular Abnormalities in Adults With Philadelphia Chromosome-Positive Acute Leukemia.", *Blood*, vol. 78, no. 9, pp. 2411–8, 1991.

[35] K. Suryanarayan, S. P. Hunger, S. Kohler, A. J. Carroll, W. Crist, M. P. Link, and M. L. Cleary, "Consistent Involvement of the BCR Gene by 9;22 Breakpoints in Pediatric Acute Leukemias.", *Blood*, vol. 77, no. 2, pp. 324–30, 1991.

[36] F. Pane, F. Frigeri, M. Sindona, L. Luciano, F. Ferrara, R. Cimino, G. Meloni, G. Saglio, F. Salvatore, and B. Rotoli, "Neutrophilic-Chronic Myeloid Leukemia: A Distinct Disease With a Specific Molecular Marker (BCR/ABL With C3/A2 Junction)", *Blood*, vol. 88, no. 7, pp. 2410–4, 1996.

[37] G. Wilson, "BCR-ABL Transcript With an e19a2 (c3a2) Junction in Classical Chronic Myeloid Leukemia.", *Blood*, vol. 89, no. 9, p. 3064, 1997.

[38] Greuber and Pendergast, "Role of ABL family kinases in cancer: from leukaemia to solid tumours.", *Nature Reviews Cancer*, vol. 13, no. 8, pp. 559–71, 2013.

[39] D. Sears, P. Luong, M. Yuan, G. Nteliopoulos, Y. Man, J. Melo, and S. Basu, "Functional phosphoproteomic analysis reveals cold-shock domain protein A to be a Bcr-Abl effector-regulating proliferation and transformation in chronic myeloid leukemia.", *Cell Death and Disease*, vol. 72, no. 1, e93, 2010.

[40] F. Toffalini and J.-B. Demoulin, "New insights into the mechanisms of hematopoietic cell transformation by activated receptor tyrosine kinases.", *Blood*, vol. 116, no. 14, pp. 2429–37, 2010.

[41] S. Medves and J. B. Demoulin, "Tyrosine kinase gene fusions in cancer: Translating mechanisms into targeted therapies", *Journal of Cellular and Molecular Medicine*, no. 2, pp. 237–48, 2012.

[42] E. L. Kwak, Y.-J. Bang, D. R. Camidge, A. T. Shaw, B. Solomon, R. G. Maki, S.-H. I. Ou, B. J. Dezube, P. A. Jänne, D. B. Costa, M. Varella-Garcia, W.-H. Kim, T. J. Lynch, P. Fidias, H. Stubbs, J. A. Engelman, L. V. Sequist, W. Tan, L. Gandhi, M. Mino-Kenudson, G. C. Wei, S. M. Shreeve, M. J. Ratain, J. Settleman, J. G. Christensen, D. A. Haber, K. Wilner, R. Salgia, G. I. Shapiro, J. W. Clark, and A. J. Iafrate, "Anaplastic Lymphoma Kinase Inhibition in NonSmall-Cell Lung Cancer.", *New England Journal of Medicine*, vol. 363, no. 18, pp. 1693–703, 2010.

[43] D. E. Gerber and J. D. Minna, "ALK Inhibition for Non-Small Cell Lung Cancer: From Discovery to Therapy in Record Time.", *Cancer Cell*, no. 6, pp. 548–51, 2010.

[44] S. Turner and D. Alexander, "Fusion tyrosine kinase-mediated signalling pathways in the transformation of haematopoietic cells.", *Leukaemia*, vol. 20, no. 4, pp. 572–82, 2006.

[45] S. Koschmieder, B. Gö, P. Zhang, J. Iwasaki-Arai, K. Akashi, J. L. Kutok, T. Dayaram, K. Geary, A. R. Green, D. G. Tenen, and C. S. Huettner, "Inducible chronic phase of myeloid leukemia with expansion of hematopoietic stem cells in a transgenic model of BCR-ABL leukemogenesis.", *Blood*, vol. 105, no. 1, pp. 324–34, 2005.

[46] S. Wong and O. N. Witte, "THE BCR-ABL Story: Bench to Bedside and Back.", *Annual review of immunology*, vol. 22, pp. 247–306, 2004.

[47] Lacronique, "A TEL-JAK2 Fusion Protein with Constitutive Kinase Activity in Human Leukemia.", *Science*, vol. 278, no. 5342, pp. 1309–12, 1997.

[48] T. R. Golub, "Oligomerization of the ABL Tyrosine Kinase by the Ets Protein TEL in Human Leukemia.", *Molecular and Cellular Biology*, vol. 16, no. 8, pp. 4107–4116, 1996.

[49] C. E. Tognon, C. D. Mackereth, A. M. Somasiri, L. P. Mcintosh, and P. H. B. Sorensen, "Mutations in the SAM Domain of the ETV6-NTRK3 Chimeric Tyrosine Kinase Block Polymerization and Transformation Activity.", *Molecular and Cellular Biology*, vol. 24, no. 11, pp. 4636–4650, 2004.

[50] C. Jousset, C. Carron, A. Boureux, C. T. Quang, M. Charon, J. Levin, O. Bernard, and J. Ghysdael, "A domain of TEL conserved in a subset of ETS proteins defines a specific oligomerization interface essential to the mitogenic properties of the TELPDGFRβ oncoprotein members of this family are implicated to date in a variety Cé cile Oury, Isabelle Dusant", *The EMBO Journal*, vol. 16, no. 1, pp. 69–82, 1997.

[51] S. Medves, L. A. Noël, C. P. Montano-Almendras, R. I. Albu, H. Schoemans, S. N. Constantinescu, and J. B. Demoulin, "Multiple oligomerization domains of KANK1-PDGFRβ are required for JAK2-independent hematopoietic cell proliferation and signaling via STAT5 and ERK.", *Haematologica*, no. 10, pp. 1406–14, 2011.

[52] H. Baumann, P. Kunapuli, E. Tracy, and J. K. Cowell, "The oncogenic fusion protein-tyrosine kinase ZNF198/fibroblast growth factor receptor-1 has signaling function comparable with interleukin-6 cytokine receptors.", *Journal of Biological Chemistry*, no. 18, pp. 16 198–208, 2003.

[53] X. Zhao, S. Ghaffari, H. Lodish, V. N. Malashkevich, and P. S. Kim, "Structure of the Bcr-Abl oncoprotein oligomerization domain.", *Nature Structural Biology*, no. 2, pp. 117–20, 2002.

[54] J. R. Mcwhirter, D. L. Galasso, and J. Y. J. Wang, "A Coiled-Coil Oligomerization Domain of Bcr Is Essential for the Transforming Function of Bcr-Abl Oncoproteins.", *Molecular and Cellular Biology*, vol. 13, no. 12, pp. 7587–7595, 1993.

[55] K. De Keersmaecker, M. Versele, J. Cools, G. Superti-Furga, and O. Hantschel, "Intrinsic differences between the catalytic properties of the oncogenic NUP214-ABL1 and BCR-ABL1 fusion protein kinases.", *Leukemia*, vol. 22, no. 12, pp. 2208–16, 2008.

[56] C. S. Kasyapa, P. Kunapuli, and J. K. Cowell, "HSPA1A is an important regulator of the stability and function of ZNF198 and its oncogenic derivative, ZNF198-FGFR1.", *Journal of Cellular Biochemistry*, no. 5, pp. 1308–17, 2007.

[57] A. T. Look, "Oncogenic Transcription Factors in the Human Acute Leukemias.", *Science*, vol. 278, no. 5340, pp. 1059–64, 1997.

[58] A. G. Tsai and M. R. Lieber, "Mechanisms of chromosomal rearrangement in the human genome.", *BMC Genomics*, vol. 11, no. 1, 2010.

[59] G. Q. Daley and D. Baltimore, "Transformation of an interleukin 3-dependent hematopoietic cell line by the chronic myelogenous leukemia-specific P21Obcr/abl protein (Abelson murine leukemia virus/Philadelphia chromosome/tyrosine kinase)", *Medical Sciences*, vol. 85, pp. 9312–9316, 1988.

[60] J. Warner, J. Wang, K. Takenaka, S. Doulatov, J. Mckenzie, L. Harrington, and J. Dick, "Direct evidence for cooperating genetic events in the leukemic transformation of normal human hematopoietic cells", *Leukemia*, vol. 19, pp. 1794–1805, 2005.

[61] A. S. Lundberg, S. H. Randell, S. A. Stewart, B. Elenbaas, K. A. Hartwell, M. W. Brooks, M. D. Fleming, J. C. Olsen, S. W. Miller, R. A. Weinberg, and W. C. Hahn, "Immortalization and transformation of primary human airway epithelial cells by gene transfer.", *Oncogene*, vol. 21, pp. 4577–4586, 2002.

[62] D. Cortez, L. Kadlec, and A. M. Pendergast, "Structural and Signaling Requirements for BCR-ABL-Mediated Transformation and Inhibition of Apoptosis.", *Molecular and Cellular Biology*, vol. 15, no. 10, pp. 5531–5541, 1995.

[63] R. Chiarle, J. Z. Gong, I. Guasparri, A. Pesci, J. Cai, J. Liu, W. J. Simmons, G. Dhall, J. Howes, R. Piva, and G. Inghirami, "NPM-ALK transgenic mice spontaneously develop T-cell lymphomas and plasma cell tumors.", *Blood*, no. 5, pp. 1919–27, 2003.

[64] M. Nieborowska-Skorska, A. Slupianek, L. Xue, Q. Zhang, P. N. Raghunath, G. Hoser, M. A. Wasik, S. W. Morris, and T. Skorski, "Role of Signal Transducer and Activator of Transcription 5 in Nucleophosmin/ Anaplastic Lymphoma Kinase-mediated Malignant Transformation of Lymphoid Cells.", *Cancer Research*, vol. 61, no. 17, pp. 6517–6523, 2001.

[65] V. Sexl, R. Piekorz, R. Moriggl, J. Rohrer, M. P. Brown, K. D. Bunting, K. Rothammer, M. F. Roussel, and J. N. Ihle, "Stat5a/b contribute to interleukin 7induced B-cell precursor expansion, but abl- and bcr/abl-induced transformation are independent of Stat5.", *Blood*, vol. 96, no. 6, pp. 2277–83, 2000.

[66] R. L. Ilaria, "Bcr/Abl, leukemogenesis, and genomic instability: A complex partnership.", *Leukemia Research*, no. 11, pp. 971–3, 2002.

[67] J. W. Voncken, C. Morris, P. Pattengale, G. Dennert, C. Kikly, J. Groffen, and N. Heisterkamp, "Clonal development and karyotype evolution during leukemogenesis of BCR/ABL transgenic mice.", *Blood*, no. 4, pp. 1029–36, 1992.

[68] H. Salloukh and P. Laneuville, "Increase in mutant frequencies in mice expressing the BCR-ABL activated tyrosine kinase.", *Leukemia*, vol. 14, no. 8, pp. 1401–1404, 2000.

[69] J. M. Brain, N. Goodyer, and P. Laneuville, "Measurement of Genomic Instability in Preleukemic P190 BCR/ABL Transgenic Mice Using Inter-Simple Sequence Repeat Polymerase Chain Reaction", *Cancer Research*, vol. 63, no. 16, pp. 4895–4898, 2003.

[70] P. Laneuville, G. Sun, M. Timm, and M. Vekemans, "Clonal Evolution in a Myeloid Cell Line Transformed to Interleukin-3 Independent Growth by Retroviral Transduction and Expression of p210bcr-abl.", *Blood*, vol. 80, no. 7, pp. 1788–97, 1992.

[71] Q. Yu, J. Brain, P. Laneuville, and D. Osmond, "Suppressed apoptosis of pre-B cells in bone marrow of pre-leukemic p190 bcr/abl transgenic mice.", *Leukemia*, vol. 15, no. 5, pp. 819–27, 2001.

[72] R. Kuribara, H. Honda, H. Matsui, T. Shinjyo, T. Inukai, K. Sugita, S. Nakazawa, H. Hirai, K. Ozawa, and T. Inaba, "Roles of Bim in Apoptosis of Normal and Bcr-Abl-Expressing Hematopoietic Progenitors.", *Molecular and Cellular Biology*, vol. 24, no. 14, pp. 6172–6183, 2004.

[73] N. Takeda, M. Shibuya, Y. Maru, and O. N. Witte, "The BCR-ABL oncoprotein potentially interacts with the xeroderma pigmentosum group B protein.", *Medical Sciences*, vol. 96, no. 1, pp. 203–207, 1999.

[74] I. Weinstein and A. Joe, "Oncogene addiction.", *Cancer Res.*, no. 9, 2008.

[75] N. Boissel, A. Renneville, V. Biggio, N. Philippe, X. Thomas, J.-M. Cayuela, C. Terre, I. Tigaud, S. Castaigne, E. Raffoux, S. De Botton, P. Fenaux, H. Dombret, and C. Preudhomme, "Prevalence, clinical profile, and prognosis of NPM mutations in AML with normal karyotype.", *Blood*, vol. 106, no. 10, pp. 3618–20, 2005.

[76] Q. Liu, J. Schwaller, J. Kutok, D. Cain, J. C. Aster, I. R. Williams, and D. G. Gilliland, "Signal transduction and transforming properties of the TEL- TRKC fusions associated with and acute myelogenous leukemia.", *EMBO Journal*, no. 8, pp. 1827–38, 2000.

[77] R. Horie, M. Watanabe, T. Ishida, T. Koiwa, S. Aizawa, K. Itoh, M. Higashihara, M. E. Kadin, and T. Watanabe, "The NPM-ALK oncoprotein abrogates CD30 signaling and constitutive NF-KB activation in anaplastic large cell lymphoma.", *Cancer Cell*, no. 4, pp. 353–64, 2004.

[78] D. Cortez, G. Reuther, and A. M. Pendergast, "The Bcr-Abl tyrosine kinase activates mitogenic signaling pathways and stimulates G1-to-S phase transition in hematopoietic cells.", *Oncogene*, vol. 15, 1997.

[79] S. Mathas, M. Hinz, I. Anagnostopoulos, D. Krappmann, A. Lietz, F. Jundt, K. Bommert, F. Mechta-Grigoriou, H. Stein, B. Dörken, and C. Scheidereit, "Aberrantly expressed c-Jun and JunB are a hallmark of Hodgkin lymphoma cells, stimulate proliferation and synergize with NF-κB.", *EMBO Journal*, no. 15, pp. 4104–13, 2002.

[80] J. H. S. Kabarowskil, P. B. Allen, and L. M. Wiedemann, "A temperature sensitive p210 BCR-ABL mutant defines the primary consequences of BCR-ABL tyrosine kinase expression in growth factor dependent cells.", *The EMBO Journal*, vol. 13, no. 24, pp. 5887–5895, 1994.

[81] T.-L. Gu, Z. Tothova, B. Scheijen, J. D. Griffin, D. G. Gilliland, and D. W. Sternberg, "NPM-ALK fusion kinase of anaplastic large-cell lymphoma regulates survival and proliferative signaling through modulation of FOXO3a.", *Blood*, vol. 103, pp. 4622–4629, 2004.

[82] Albert-László Barabási, "Network Medicine From Obesity to the Diseasome", *New England Journal of Medicine*, vol. 357, no. 4, pp. 404–7, 2007.

[83] T. Pawson and R. Linding, "Network medicine.", *FEBS Letters*, no. 8, pp. 1266–70, 2008.

[84] H. Blume-Jensen, "Oncogenic kinase signalling.", *Nature*, vol. 411, no. 6835, pp. 355–65, 2001.

[85] S. Reckel, R. Hamelin, S. Georgeon, F. Armand, Q. Jolliet, D. Chiappe, M. Moniatte, and O. Hantschel, "Differential signaling networks of BcrAbl p210 and p190 kinases in leukemia cells defined by functional proteomics.", *Leukemia*, no. 7, pp. 1502–1512, 2017.

[86] J. Cutler, R. Tahir, S. Sreenivasamurthy, C. Mitchell, S. Renuse, R. Nirujogi, A. Patil, M. Heydarian, X. Wong, X. Wu, T.-C. Huang, M.-S. Kim, K. Reddy, and A. Pandey, "Differential signaling through p190 and p210 BCR-ABL fusion proteins revealed by interactome and phosphoproteome analysis.", *Leukemia*, no. 7, pp. 1513–1524, 2017.

[87] O. Hantschel and G. Superti-Furga, "Regulation of the c-Abl and Bcr-Abl tyrosine kinases.", *Nature Reviews*, vol. 5, no. 1, pp. 33–44, 2004.

[88] G. E. Winter, U. Rix, S. M. Carlson, K. V. Gleixner, F. Grebien, M. Gridling, A. C. Muller, F. P. Breitwieser, M. Bilban, J. Colinge, P. Valent, K. L. Bennett, F. M. White, and G. Superti-Furga, "Systems-pharmacology dissection of a drug synergy in imatinib- resistant CML.", *Nature Chemical Biology*, vol. 8, no. 11, pp. 905–912, 2012.

[89] J. Li, U. Rix, B. Fang, Y. Bai, A. Edwards, J. Colinge, K. L. Bennett, J. Gao, L. Song, S. Eschrich, G. Superti-Furga, J. Koomen, and E. B. Haura, "A chemical and phosphoproteomic characterization of dasatinib action in lung cancer.", *Nature Chemical Biology*, vol. 6, no. 4, pp. 291–9, 2010.

[90] H. Steen, M. Fernandez, S. Ghaffari, A. Pandey, and M. Mann, "Phosphotyrosine Mapping in Bcr/Abl Oncoprotein Using Phosphotyrosine-specific Immonium Ion Scanning", *Molecular & Cellular Proteomics*, vol. 2, pp. 138–145, 2003.

[91] A. R. Salomon, S. B. Ficarro, L. M. Brill, A. Brinker, Q. T. Phung, C. Ericson, K. Sauer, A. Brock, D. M. Horn, P. G. Schultz, E. C. Peters, and C.-H. Wong, "Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry", *Proc Natl Acad Sci U S A.*, pp. 443–8, 2003.

[92] P. Giansanti, C. Preisinger, K. V. M. Huber, M. Gridling, G. Superti-Furga, K. L. Bennett, and A. J. R. Heck, "Evaluating the promiscuous nature of tyrosine kinase inhibitors assessed in A431 epidermoid carcinoma cells by both chemical- and phosphoproteomics.", *ACS Chemical Biology*, no. 7, pp. 1490–8, 2014.

[93] B. J. Skaggs, M. E. Gorre, A. Ryvkin, M. R. Burgess, Y. Xie, Y. Han, E. Komisopoulou, L. M. Brown, J. A. Loo, E. M. Landaw, C. L. Sawyers, T. G. Graeber, and J. Kuriyan, "Phosphorylation of the ATP-binding loop directs oncogenicity of drug-resistant BCR-ABL mutants.", *Proc Natl Acad Sci U S A.*, vol. 103, no. 51, pp. 19 466–71, 2006.

[94] T. G. Graeber, J. R. Heath, B. J. Skaggs, M. E. Phelps, F. Remacle, and R. D. Levine, "Maximal entropy inference of oncogenicity from phosphorylation signaling", *Proc Natl Acad Sci U S A.*, vol. 107, no. 13, pp. 6112–7, 2010.

[95] K. De Keersmaecker, M. Porcu, L. Cox, T. Girardi, R. Vandepoel, J. O. De Beeck, O. Gielen, N. Mentens, K. L. Bennett, and O. Hantschel, "NUP214-ABL1-mediated cell proliferation in T-cell acute lymphoblastic leukemia is dependent on the LCK kinase and various interacting proteins.", *Haematologica*, no. 1, pp. 85–93, 2014.

[96] U. Rix, L. R. Rix, A. Terker, N. Fernbach, O. Hantschel, M. Planyavsky, F. Breitwieser, H. Herrmann, J. Colinge, K. Bennett, M. Augustin, J. Till, M. Heinrich, P. Valent, and G. Superti-Furga, "A comprehensive target selectivity survey of the BCR-ABL kinase inhibitor INNO-406 by kinase profiling and chemical proteomics in chronic myeloid leukemia cells.", *Leukemia*, vol. 24, no. 1, pp. 44–50, 2009.

[97] D. M. Weinstock, E. Brunet, and M. Jasin, "Induction of chromosomal translocations in mouse and human cells using site-specific endonucleases.", in *Journal of the National Cancer Institute - Monographs*, 2008, pp. 20–4.

[98] E. H. Breese, C. Buechele, C. Dawson, M. L. Cleary, and M. H. Porteus, "Use of genome engineering to create patient specific MLL translocations in primary human hematopoietic stem and progenitor cells.", *PLoS ONE*, no. 9, e0136644, 2015.

[99] M. Piganeau, H. Ghezraoui, A. De Cian, L. Guittat, M. Tomishima, L. Perrouault, O. René, G. E. Katibah, L. Zhang, M. C. Holmes, Y. Doyon, J.-P. Concordet, C. Giovannangeli, M. Jasin, and E. Brunet, "Cancer translocations in human cells induced by zinc finger and TALE nucleases.", *Genome Research*, no. 7, pp. 1182–93, 2103.

[100] P. S. Choi and M. Meyerson, "Targeted genomic rearrangements using CRISPR/Cas technology.", *Nature Communications*, vol. 5, 2014.

[101] R. Torres, M. C. Martin, A. Garcia, J. C. Cigudosa, J. C. Ramirez, and S. Rodriguez-Perales, "Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPRCas9 system.", *Nature Communications*, vol. 5, no. 3964, 2014.

[102] D. Maddalo, E. Manchado, C. P. Concepcion, C. Bonetti, J. A. Vidigal, Y.-C. Han, P. Ogrodowski, A. Crippa, N. Rekhtman, E. De Stanchina, S. W. Lowe, and A. Ventura, "In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system.", *Nature*, vol. 516, no. 7531, pp. 423–7, 2014.

[103] T. Wang, J. J. Wei, D. M. Sabatini, E. S. Lander, T. Wang, J. J. Wei, D. M. Sabatini, and E. S. Lander, "Genetic Screens in Human Cells Using the CRISPR-Cas9 System Terms of Use Genetic screens in human cells using the CRISPR/Cas9 system", *Science*, vol. 343, no. 6166, pp. 80–84, 2013.

[104] O. T. Schubert, H. L. Röst, B. C. Collins, G. Rosenberger, and R. Aebersold, "Quantitative proteomics: challenges and opportunities in basic and applied research.", *Nature Protocols*, no. 7, pp. 1289–1294, 2017.

[105] T. Geiger, J. Wisniewski, J. Cox, S. Zanivan, M. Kruger, Y. Ishihama, and M. Mann, "Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics.", *Nature protocols*, vol. 6, no. 2, pp. 147–57, 2011.

[106] N. Rauniyar and J. R. I. Yates, "Isobaric labeling-based relative quantification in shotgun proteomics.", *Journal of Proteome Research*, vol. 13, no. 12, pp. 5293–309, Dec. 2014.

[107] H. Wang, S. Alvarez, and L. M. Hicks, "Comprehensive Comparison of iTRAQ and Label-free LC-Based Quantitative Proteomics Approaches Using Two Chlamydomonas reinhardtii Strains of Interest for Biofuels Engineering", *Journal of Proteome Research*, vol. 11, pp. 487–501, 2012.

[108] J. V. Olsen, M. Vermeulen, A. Santamaria, C. Kumar, M. L. Miller, L. J. Jensen, F. Gnad, J. Cox, T. S. Jensen, E. a. Nigg, S. Brunak, and M. Mann, "Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis.", *Science Signaling*, vol. 3, no. 104, ra3, Jan. 2010.

[109] Y. Zhang, A. Wolf-Yadlin, P. L. Ross, D. J. Pappin, J. Rush, D. A. Lauffenburger, and F. M. White, "Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules", *Molecular & Cellular Proteomics*, vol. 4, pp. 1240–1250, 2005.

[110] P. Mertins, D. R. Mani, K. V. Ruggles, M. A. Gillette, K. R. Clauser, P. Wang, X. Wang, J. W. Qiao, S. Cao, F. Petralia, E. Kawaler, F. Mundt, K. Krug, Z. Tu, J. T. Lei, M. L. Gatza, M. Wilkerson, C. M. Perou, V. Yellapantula, K.-L. Huang, C. Lin, M. D. Mclellan, P. Yan, S. R. Davies, R. R. Townsend, S. J. Skates, J. Wang, B. Zhang, C. R. Kinsinger, M. Mesri, H. Rodriguez, L. Ding, A. G. Paulovich, D. Fenyo, M. J. Ellis, S. A. Carr, and N. Cptac, "Proteogenomics connects somatic mutations to signaling in breast cancer", *Nature*, vol. 534, no. 7605, pp. 55–62, 2016.

[111] E. S. Boja and H. Rodriguez, "Proteogenomic convergence for understanding cancer pathways and networks", *Clinical Proteomics*, vol. 11, no. 22, p. 22, 2014.

Chapter 2

Harnessing gene-editing techniques for the generation of chromosomal aberrations

This chapter provides a historical overview of the gene editing technologies and how they furthered research in the life sciences, including the latest addition to the genetic engineering arsenal: CRISPR-Cas9. It is concluded by an introduction to chromosomal rearrangements and their genesis using nuclease-based gene editing.

2.1 Precision Genome Editing for Systems Biology - A Temporal Perspective

Originally published as : Voellmy, F. and Linding, R. (2017) Precision Genome Editing for Systems Biology - A Temporal Perspective, in Systems Biology (eds J. Nielsen and S. Hohmann), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

Over the past 60 years, the development of techniques allowing the production and manipulation of DNA has provided molecular biologists with the tools to make many of the major advances impacting basic science, biotechnology, and medicine. The discovery that endogenous nucleases could be exploited for their sequence-specific DNA-cleaving abilities within the cell or organism itself allowed biologists to probe the effects of gene modification, which is essential to the understanding of gene function. In this chapter we will cover the emergence and development of the gene-editing era over time beginning with zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), and continue with the newest popular technique of clustered regularly interspaced short palindromic repeat (CRISPR), which, since its demonstration to be applicable as a DNA editing platform in 2013, has already been adopted by many research groups worldwide. The ease of design of this technique has rendered it amenable to high-throughput studies, such as genome-wide loss-of-function screens in mammalian cells and, as a consequence, also to systems biology. As the technique establishes itself and becomes potent for clinical applications such as gene therapy, nuclease specificity and off-target identification must be thoroughly examined. We will therefore also cover important considerations for the design of nuclease-based gene-editing research projects. Finally, we will focus on the study of chromosomal rearrangements that can be induced by means of two DNA double-strand breaks (DSB)s using sequence-specific nucleases, and thereby

present an interesting application of the CRISPR-Cas method for systems biology.

2.2 Early Techniques in DNA Alterations

The advent of recombinant DNA technology in the 1970s was at the origin of the gene modification trend using enzymes purified from bacteria to cut and ligate DNA, and further using the bacteria as a host to produce the DNA of interest in large quantities, allowing the targeted alteration of DNA molecules. A gene cloned using this technology could then be introduced into living cells via homologous recombination [1]. Several studies have reported the possibility of transferring a modified, cloned sequence to the germline of a living organism by genetically manipulating murine embryonic stem cells established in culture. When reintroduced into mouse blastocysts, these cells were capable of colonizing germ cell lineages, and the introduced genes, for example conferring antibiotic resistance [2] or essential to cellular building block production [3], were transmitted into the transgenic generation. These studies underlined the vast potential of gene therapy, notably to correct genes carrying a malignant mutation. Moreover, the advantage of performing site-specific genetic manipulation without having to extract the DNA from its genomic context was apparent. Work carried out in bacteria and yeast to understand natural DNA repair pathways showed that cells possess the machinery to repair dangerous DSBs in their DNA [4], and suggested that determining methods to create precise breaks, along with recombinant DNA technology, would be critical for genome editing. Early techniques to create DSBs already made use of DNA base pair recognition by oligonucleotides or small molecules. Molecules capable of recognizing DNA sequences were coupled to cleavage reagents such as bleomycin, thereby creating a system for site-targeted modification. Although not applicable on larger scales, these studies underscored the idea of using base-pairing to a host's DNA as a means for site-specific binding [5]–[8]. One example of an early technique using nucleases, which is based on base-pair recognition, is self-splicing introns capable of modifying DNA or RNA sequences. Yeast mobile group II introns such as aI1 and aI2 encode reverse transcriptases that convert intron RNAs into complementary DNA (cDNA) and insert the intron at a high frequency at the same site in an intronless allele [9], [10], also known as "homing." To do so, an intron-encoded endonuclease acts as a ribozyme, with the excised RNA creating a break in the sense strand and the protein moiety in the antisense strand of the double-stranded DNA at a precise location, allowing the insertion of the intron into the genomic DNA [11]. Interestingly, although the aI1 and aI2 self-splicing introns are closely related, their respective insertion sites are different, suggesting endonuclease specificity [12]. Similarly, based on a self- splicing intron discovered to mediate trans-splicing of oligonucleotides, a targeted RNA-cleaving ribozyme was demonstrated to be suitable for altering the sequence of specific transcripts [13]. By fusing the domains of two existing endonucleases, a highly sequence-specific artificial endonuclease was created [14]. Although the methods introduced up to this point make use of Watson–Crick base pairing in to recognize specific genetic sequences, they lack flexibility and scalability. The following technologies, namely zinc-finger nucleases, transcription activator-like effector (TALE) nucleases, and most

recently CRISPR-Cas systems, build on the use of base pairing but are sufficiently modulable to – in theory – recognize any DNA sequence occurring in complex genomes.

2.3 Zinc-Finger Nucleases

The zinc finger is a β β α domain with a zinc ion coordinated by two cysteine and two histidine residues, recognizing three consecutive base pairs [15]. Zinc-finger domains [16], [17] can be tethered to the sequence-independent DNA cleaving domain of the bacterial restriction endonuclease Fok1 [18], thereby creating a site-specific nuclease where the target sequence recognition is dictated by the DNA-binding domain (DBD) of the zinc-finger protein. Zinc-finger domains are often found as tandem repeats, permitting the recognition of asymmetric sequences. Structural advances in the field have enabled the generation of synthetic arrays of over three zinc-finger domains by modifying the linker between repeats, allowing virtually any 18-bp sequence to be targeted [19], a length sufficient to be unique even in a genome as complex as Homo sapiens' [15]. The approach was successfully demonstrated to introduce mutations in Drosophila as well as mammalian cells [20], [21], and it can be used to introduce different genomic alterations such as point mutations, deletions, insertions, inversion, duplications, and translocations [22]. It nevertheless remains difficult to adapt to large-scale applications where many sequences are targeted, because of tedious protein design requiring many steps (indeed, clever design alone is not sufficient, and selection strategies must be applied to optimize the zinc finger for a target) [19].

2.4 TALENs

A similar approach was applied to generate an improved version of naturally occurring bacterial TALEs, by fusing the TALE DBD to various other functional domains. TALEs are naturally encoded by the plant pathogenic bacteria *Xanthomonas spp.*, and when injected into host cells have the capacity to alter transcription by binding to genomic DNA, thereby increasing pathogenicity. These proteins are comprised of a series of 33–35 amino acid repeats, each recognizing 1bp. For genome editing purposes, arrays of highly conserved repeats derived from the DBD of TALEs can be customized to recognize specific DNA motifs. Within each repeat of an array, two amino acid residues named repeat variable diresidues (RVD)s [23] are responsible for specificity in binding of a single nucleotide [24], and the assembly of repeats with selected RVDs determines the target sequence and its length. The unraveling of this RVD-to- protein code [24], [25] enabled researchers to engineer the DBD to target any sequence of interest (typically 15–20bp [26]–[28]) and thereby customize this system with more ease than previous methods. However, this means that a new DBD needs to be engineered for each 15 – 20 bp target, requiring significant time and effort. Indeed, extensive identical repeats can present synthesis challenges, and methods have been developed to work around this limitation [28], [29]. TALE DBDs can be combined with a Krüppel-associated box (KRAB) domain to achieve gene silencing by targeting the KRAB to transcription start sites. Transcription is thereby repressed in the nucleus, rather than being degraded post-transcriptionally

as is the case with RNAi. In order to induce DSBs, TALENs are created by fusing the Fok1 nuclease effector domain to the DBD of TAL (transcription activator-like) [30]. Since Fok1 requires dimerization for activity, this imposes a requirement to have a pair of TALEN for the correct positioning of Fok1 domains with respect to the desired cut site. This, in turn, strongly increases specificity of directed DSB [31], and off-target effects are typically low [27], [32], [33].

Despite the aforementioned difficulties in TALE design, a study by Kim *et al.* [34] targeted over 18 000 genes by constructing a TALEN library, thereby demonstrating the potential of these nucleases to large-scale studies.

2.5 CRISPR-Cas9

The most recent addition to the programmable nuclease toolbox is the CRISPR-Cas system. In 2007 [35], it was shown that the CRISPR-Cas system was able to confer adaptive immunity [36]–[38] to *Streptococcus thermophilus* bacteria against phage infection. From an industrial viewpoint, the ability to aid the workhorse bacterial starter cultures in overcoming phage exposure remains an area of application that is of great financial interest [39]. Briefly, this system, which is widespread in bacteria and archaea, allows the incorporation of short fragments of DNA from invading viruses or plasmids into the organism's own genome, and upon future encounters with the same virus or plasmid family, the latter threat can be decimated by targeting DNA matching the previously incorporated sequences [40] (Fig. 2.1). This antiviral defense is the result of a series of events. First, a short fragment of the foreign DNA is incorporated into the CRISPR array, which is an array of identical repeats interspersed with invader DNA-targeting spacers of approximately 20 bp, called protospacers. These spacers are transcribed as precursor CRISPR RNA (pre-crRNA) and are then further processed to become individual CRISPR RNAs (crRNAs), serving as guides for Cas proteins to interfere with invading DNA. The CRISPR-Cas loci also transcribe an operon of cas genes. The encoded Cas protein components then cleave the invading DNA at locations complementary to the crRNA sequence. Several CRISPR-Cas systems have been found to follow this overall scheme, with differences in the exact mechanisms governing site recognition and cleavage, and have been classified into three main types I, II, and III, along with subtypes for each [41], [42]. The system types differ in the expressed Cas proteins and their mode of action. The type II system has proven to be most useful for genome editing purposes because of the fact that it requires only a single Cas protein, Cas9 (previously known as Cas5, Csn1, or Csx12), to mediate both sequence specificity and nuclease activity [43]–[45].

In contrast, type I and III systems rely on a large multimeric crRNA – Cas ribonucleoprotein complex, which results in a setup of greater complexity is and thus less amenable to development as a gene-editing tool. Trans-activating crRNA (tracrRNA) is a small noncoding RNA that is complementary to the repeated sequences in the pre-crRNA, and is utilized by the type II CRISPR-Cas system in *Streptococccus pyogenes* for pre-crRNA processing by binding to the crRNA and triggering cleavage by RNase III (in presence of Cas proteins), and has also been reported to be necessary for nuclease sequence specificity. Jinek *et al.* then demonstrated that the final

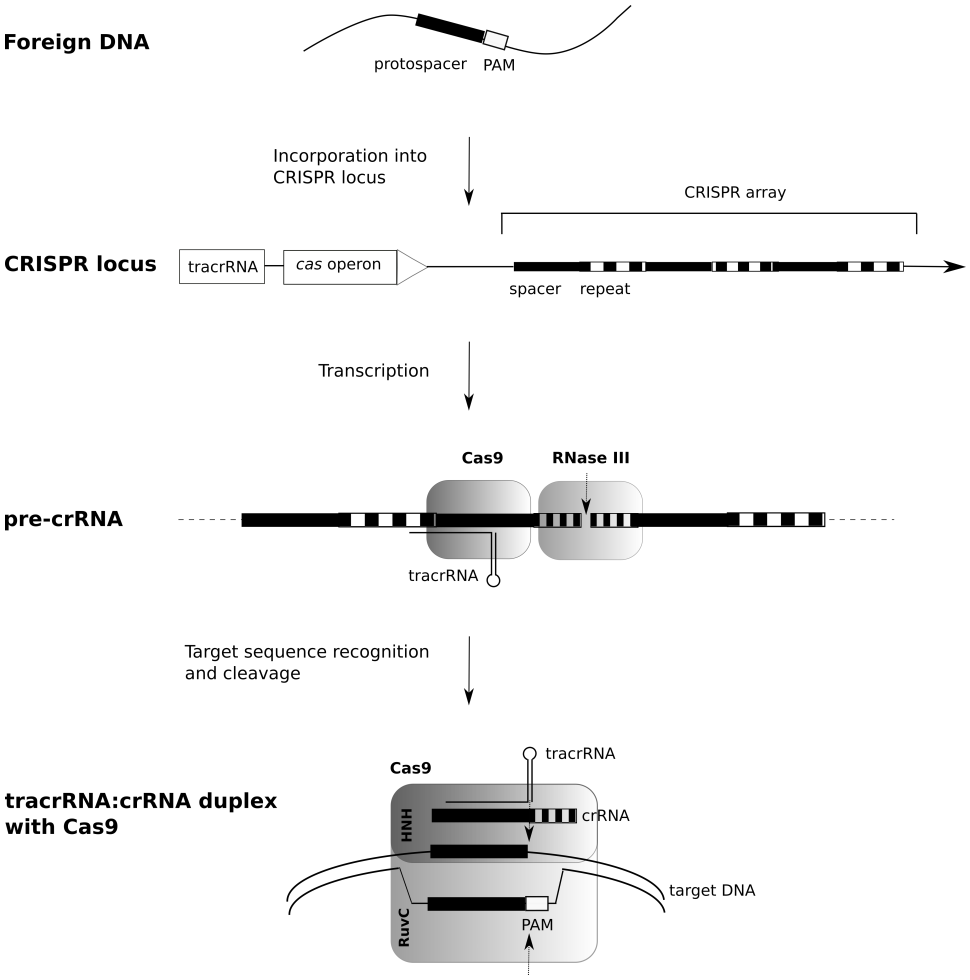


FIGURE 2.1: Schematic of the type II CRISPR- Cas9 mechanism of adaptive immunity. Invading DNA is incorporated into a CRISPR array, alternating between repeated short fragments of foreign DNA of approximately 20bp (protospacer) and identical repeats (PAM). The CRISPR locus is transcribed into pre-crRNA and finally, after binding of tracrRNA to pre-crRNA, matures into crRNA. crRNA and tracrRNA form a duplex triggering cleavage of target DNA by the Cas9 nuclease.

CRISPR-Cas9 complex was a dual-RNA-guided DNA endonuclease in which a tracrRNA:crRNA duplex is formed and bound to Cas9, directing it to a target with complementarity to the crRNA [43].

2.6 CRISPR-Cas9

The protospacer adjacent motif (PAM) is an essential component of the type II system and is found flanking the 3' end of the target site, for example, 5' NGG for *S. thermophilus* and *S. pyogenes*. The critical nature of the PAM was shown by PAM mutations in phage genomes interfering with the antiviral defense mechanism [46]. The RNA–DNA duplex is formed first at the PAM site and then allows interrogation for upstream complementarity. In support of this, catalytically inactive Cas9 mutants were found to bind at

many more sites than the expected exact sequence matches. These sites were typically adjacent to a PAM and presented partial complementarity to the guide sequence. However, active Cas9 infrequently cleaves at these off-target sites, thereby pointing toward uncoupled DNA recognition and cleaving [47]. The family of Cas9 proteins possesses two signature nuclease domains: HNH and RuvC. When there is complementarity between the crRNA and the guide sequence, the Cas9 HNH domain cleaves the complement to the 20-nt sequence of the crRNA, whereas the RuvC domain cleaves the noncomplementary strand. Nickases that create a single break rather than a DSB can be engineered by mutating one of the two domains depending on the strand where cleavage is desired [43], [44]. Elucidation of the crystal structure of *S. pyogenes* [48], [49] showed that, when dissociated from guiding elements, the Cas9 protein remains in an auto-inhibited conformation in which the HNH domain active site is blocked by the RuvC domain. Further simplifying the practical application of the CRISPR-Cas9, the tracr- RNA:crRNA complex was found to be replaceable by a single guide RNA (sgRNA) of 20 bp [43]. This short RNA is essentially a combination of the tracrRNA and crRNA, where the guiding properties of the crRNA and the secondary structure of the duplex required for DNA cleavage are maintained. This allowed a system where the simple change of the sgRNA could re-target Cas9 to any sequence adjacent to a PAM. The implementation of the CRISPR-Cas9 as a gene editing tool was shown in several studies [50]–[53], as well as the possibility of "multiplexing" where multiple sgRNAs are used for targeting [50], [51]. Moreover, similar to fusion TALE DBDs with KRAB domains, mutant Cas9 can be fused to effector domains such as KRAB to induce transcriptional repression.

2.7 Considerations of Gene-Editing Nuclease Technologies

2.7.1 Repairing Nuclease-Induced DNA Damage

After the DNA DSB created by one of these methods, different repair pathways may be stimulated in mammalian cells: the error-prone non-homologous end-joining (NHEJ), or the high-fidelity homology-directed repair (HDR) [54]. These repair mechanisms can be taken advantage of to introduce targeted alterations in the DNA sequence [22]. The HDR pathway is critical in maintaining genetic stability and preventing cancer-related genetic alterations, and can be activated upon DSB if a donor template is provided via an unbroken sister chromatid or homologous chromosome, or supplied by an external source. The HDR pathway, which is less error-prone than NHEJ, when presented with a template double-stranded DNA sequence partially overlapping with the region surrounding the DSB, can lead to the precise introduction of mutations or entire sequences up to 7.6 kb [55]. Alternatively, NHEJ mediates inaccurate repair of the DSB, often resulting in substitutions at the target site as well as insertions or deletions (indels) [20], [56], [57]. In the case where the DSB site is within a coding region of a gene, the NHEJ repair machinery promotes frameshift mutations, potentially resulting in the knock-out of gene function [58]–[60]. Indels introduced by NHEJ after a DSB can vary from a few nucleotides to tens of nucleotides [61]. Large deletions

(several megabases) [62] as well as inversions [63] can be induced by cutting at two distinct sites on the same chromosome, whereas translocations [64]–[68] can be created with DSBs on different chromosomes. The type of DNA break can be used to influence whether HDR or NHEJ will mediate the repair. Indeed, nicks (i.e., single-strand breaks rather than DSBs) can stimulate the involvement of HDR [50], [51], [69]–[71]. Specific insertions are mediated by supplying cells with a double-stranded template, making use of the HDR pathway. Despite this, activation of NHEJ can still lead to undesired alteration of the targeted genomic site. Development of methods that favor HDR over NHEJ will be important to reduce unwanted adverse effects, and also for systems employing the Cas9 nuclease for site-specific introduction of new genetic material.

2.7.2 Nuclease Specificity

In order for programmable nucleases to be amenable to use in a clinical setting, as well as to be truly relevant to biological research in the context of complex genomes, their exact mechanism of action needs to be determined in order to confidently predict and assess potential off-target cleavage activity. Recognizing that all nucleases carry imperfections to some degree and addressing the question of target specificity are especially critical, given the permanent nature of these modifications. This should be a concern both when choosing guide sequences for a desired target and during the investigation of off-targets as a "by-product" of nuclease activity. The first point can be addressed using several online tools that have been developed to aid researchers in their choice of target sites for a particular CRISPR-Cas application aiming to achieve a high "on" rate; they have recently been reviewed and compared [72], [73]. Some examples include sgRNA Designer and sgRNA Scorer [74], [75], which Lee *et al.* [73] have shown to have poor overlapping top ranking hits. Cross-checking the sites given by different prediction algorithms could be one way of selecting top targets. Experimental validation of nuclease on-target activity can be carried out in several ways. Most methods detect indels generated by DSB repair by the NHEJ pathway, either using enzymes and separating the cleavage products on an agarose gel [76], or by Sanger sequencing of clones [61], [77] or the entire population [78]. Hendel *et al.* [61] developed a method called single-molecule real-time (SMRT) sequencing, which is capable of detecting both HDR and NHEJ repair following DNA DSB. An alternative is the BLESS (direct in situ breaks labeling, enrichment on streptavidin, and next-generation sequencing) technique, which labels DSBs and, following enrichment, makes use of next-generation sequencing to directly measure DSB levels [79]. Assessment of the extent of off-targets of a gene-editing nuclease system can be performed by identifying expected at-risk sites and monitoring these for undesired cleavage events. This is rendered difficult because of large genome size and the large number of potential cleavage sites, in addition to cleavage of identical sites depending on genomic context [80]. Thorough determination of the extent of nuclease promiscuity could be achieved by less biased methods, for example, whole-genome sequencing [81]. In the case of ZFNs and TALENs, structure- and selection-based approaches [82], [83] have been developed to improve specificity and reduce toxicity, and several assays have been developed to determine their specificity [81], [84]. These are typically applied

to individual nuclease systems, as a systematic evaluation remains elusive because of the difficulty in generating proteins with varying sequence specificity. In the case of the CRISPR-Cas9 system, target recognition is imposed by base-pairing interactions between the guide RNA with its target sequence as well as recognition of the adjacent PAM sequence, and is therefore potentially well suited to probe the effects of mismatching at genome-wide levels [52], [72], [85], [86]. These studies have shown that the extent of off-site targeting, defined as the tolerance of Cas9 to mismatches in the RNA guide sequence, is dependent on the number, position, and distribution of mismatches spanning the guide sequence as opposed to previous suggestions of a 8–12 nucleotide specificity-determining seed region [43], [50]. Deviations in the PAM sequences are less well tolerated in comparison to mismatches within the target sequence, although Cas9 recognizes alternative PAMs to the optimal NGG: NAG, NGA, NCG, NGC, NGT, NTG, and NAA [72], [87]. Selecting a Cas9 ortholog derived from a different species may provide higher binding specificities [88]. Off-target sites can be organized into three groups depending on their differences with the true target sequence: sequences of same length but containing base mismatches; shorter sequences with missing bases; and longer sequences containing more bases [89]. More recent studies have shown that CRISPR-Cas systems display significant nonspecific nuclease activity at off-target sites containing several mismatches with the desired target [52], [72], [85], [86], [90], thus underscoring the importance of developing methods to globally assess off-target activity.The consequences of these mismatches may vary depending on the application of interest using Cas9, given its decoupled target recognition and cleavage activities. Indeed, although Cas9 may be tolerant to a number of mismatches

2.8 Considerations of Gene-Editing Nuclease Technologies

throughout the length of the guide sequence in terms of recognition, its cleavage activity is more conservative [91]. Ran *et al.* have shown that by employing an approach based on pairs of offset guide RNAs and a mutant Cas9 nickase, specificity can be increased by up to 1500 times that of wild-type Cas9 [92], [93]. This is achieved by the two nicks at the target site effectively resembling a DSB, thereby promoting indel formation, whereas any off-target effects from the nickase are precisely repaired. It should be noted that this increase in specificity might in large part be due to the fact that a pair of Cas9 nickases recognizes double the length of a single sgRNA, that is, 40 nt rather than 20 nt, which one would expect to have a significant impact on selectivity. Moreover, Cas9 nickases are not under the requirement to dimerize (as is the case for zinc-finger nuclease systems), and off-targets may still be generated for the individual sgRNA [87]. Introducing sgRNAs truncated by two or three nucleotides can increase on-target activity, and it is hypothesized to do so by increasing sensitivity to mismatches [94]. Another way to circumvent off-target effects as well as the mentioned shortcomings of Cas9 nickases is by fusing a mutant Cas9 to Fok1, where two of these monomers bind different target sites simultaneously [32], [87].A number of online tools exist for the prediction of off-targets, although the accuracy with which alternative cleavage sites are inferred is unclear [73]. Tsai *et al.* [87] compared the

ranking of off-target predictions generated by two tools, CRISPR Design [92] and E-CRISP [95], both based on the number and position of mismatches, with experimentally determined off-target sites. Both tools were found to have missed many of the true positive off-target sites. An improved tool, COSMID [80], takes into account partially matched sequences containing base mismatches, as is currently the case, as well as base insertions and deletions. Other resources such as Cas-OFFinder provide similar options [96]. A better understanding of target site accessibility is expected to improve the available prediction algorithms. In order to comprehensively survey the off-target space, the best strategy would be to pool the results of several of the above-mentioned resources. The large number of off-targets that are suggested by this more extensive analysis will require experimental confirmation. Methods to investigate this are required to be as unbiased as possible, as well as high throughput, in order to cover the entire genome and all the expected off-target sequences in a time- and cost-effective manner. Recently developed methods focus on the detection of DSBs to determine sites affected by nuclease activity. A Cas9 mutant possessing intact recognition but without the capacity to cleave was used to profile sites of DNA binding by ChIP-seq [91], [97]. However, this strategy was shown to be problematic because of the lack of necessary dependence between site recognition and cleavage [87]. The Digenome-seq method alleviates these issues by digesting intact complexes of DNA with the bound Cas protein followed by whole-genome sequencing [98]. The digests that are true off-sites possess a specific 5' end that can be computationally identified, and the method was shown to be comparable in sensitivity to targeted deep sequencing. A further technique, GUIDE-seq [87], integrates a short double-stranded oligonucleotide (dsODN) as a tag at DNA DSBs by NHEJ-mediated capture. These sites are then subjected to unbiased amplification and next-generation sequencing. Interestingly, in addition to detecting known and many unknown off- target sites for several different sgRNAs, the study also uncovered the existence of genomic hotspots, where the "dsODN" tag was incorporated independent of a nuclease. Moreover, the search for indels at DSBs led to the identification that chromosomal rearrangements (translocations, inversions, or large deletions) could occur at those locations, with or without the help of a nuclease.

2.9 Applications

Gene editing aided by nucleases can be used for a variety of purposes such as knock-down or site-specific mutagenesis. Zinc-finger nucleases and TALENs have been used to generate gene knock-downs and introduce mutations. Their versatility is enhanced when tethered to complementary effector domains, such as, for instance, a fluorescent marker for improved visualization of protein expression, distribution, and interactions, thus providing alternatives to techniques such as fluorescence in situ hybridization (FISH) [99], [100]. TALE nucleases can be fused to KRAB or Fok1 domains to mediate gene silencing akin to RNAi, or to create DSBs, respectively. Furthermore, when coupled to epigenetic modifying enzymes, the epigenetic landscape of cells could be probed. Although large-scale studies using TALENs are feasible [34], for the most part the major advances have been made at a far smaller scale. It is the ease of design of the latest breakthrough in gene editing,

namely the CRISPR-Cas system, that holds the promise of rendering nucleases amenable to large-scale systems applications and studies, and on which we will focus in the next sections. By combining the CRISPR-Cas9 system with various effector domains, researchers have achieved gene silencing clustered regularly interspaced short palindromic repeat interference (CRISPRi) [101]), activated gene expression, introduced new sequence elements, or manipulated Cas9 to act as a RNA-guided homing entity [52]. Experimentally, the genes needed for the CRISPR-Cas9 platform can be introduced into cells by means of viral vectors such as lentivirus and adeno-associated virus, chemical transfection, and others. The use of viral vectors is the preferred method so far, as they are well characterized and easy to apply. Practical issues pertaining to the size of the Cas9 gene can be circumvented by shorter Cas9 orthologs [102], [103]. As discussed previously, the era of nuclease-based gene editing has spurred a lot of interest in further developing the tools to reduce their bias and increase specificity. Indeed, this applies to studies using "omics" technologies to measure global systems responses to perturbation such as gene knock-out, but also to in vivo applications where the complexity of the mammalian host renders any interference an inherently systems-level problem and should be investigated as such (discussed in Section 14.6.4).

2.9.1 CRISPR Nuclease Genome-Wide Loss-of-Function Screens (CRISPRn)

An application of the CRISPR-Cas technology at a scaled level is to make use of its nuclease functionality: clustered regularly interspaced short palindromic repeat nuclease (CRISPRn). In this setup, a systematic pooled loss-of-function genetic screen is carried out by using a genome-wide sgRNA library to disrupt gene activity for all genes individually [104], [105]. While RNAi has been widely used for knock-down screens, this setup offers several advantages, the major ones being that, since repression occurs at the DNA level, it is possible (i) to obtain complete loss of function and (ii) to include non-transcribed elements in the screen such as promoters, enhancers, and introns. To perform a genome-wide CRISPRn screen, a cell line stably expressing the Cas9 nuclease is generated. A library of typically 3–5 sgRNAs per gene is created, targeting approximately 20000 genes in total, and meets criteria to minimize off-target effects while decreasing the chance that some genes cannot be probed because of a problematic sgRNA. This library is then delivered to the Cas9-expressing cells, and individual sgRNAs are used as markers to ensure a single integration per cell by sequencing. Moreover, this also acts as a barcode, providing a means to identify cells in downstream applications. Once knock-down is initiated, cells can be subjected to various conditions such as positive (resistance to drug) or negative selection [104]–[106], depending on the biological question at hand. To date, the latter ranges from defining an essential set of human genes to studying tumor growth and metastasis [107] or the innate immune response [108]. An additional example of CRISPRn is screens aiming at determining the core set of essential genes for mammalian cells [109], [110]. By screening the entire genome within multiple cell types, researchers gain a systematic overview of genes involved in essential processes; also, they can attempt to determine patterns emerging from biological systems. In these screens, the CRISPR technology is used as a means to redefine the genes that are essential to

all human cells and those that are essential depending on the context, that is, cell-type-specific. While the core-fitness genes provide a list of essential genes for further studies, the context-dependent essential genes yield interesting insights into the differences between human cells [110]. Future research studying emerging similarities and differences between cells using orthogonal methods, such as transcriptomics and proteomics, will yield an additional layer of insight as to how core genes interact to carry out the cell's critical processes. Despite small differences in study design, the different cells interrogated along with the different selection methods show the feasibility of utilizing the CRISPR-Cas platform as a novel strategy for genome-wide screens. The diversity of cell lines and cells that typically possess a limited ex vivo timespan [111], such as primary cells, demonstrates that the method is applicable in a wide range of contexts, which would provide researchers with increased possibilities to verify results. Indeed, these studies demonstrate good knock-out efficiency and high targeting specificity. The top hits are crossvalidated with results from previous screens, paving the way for future mammalian functional genomics studies investigating a multitude of perturbations and phenotypes. Nevertheless, improvements are still expected, as the current trend of targeting Cas9 to the 5' exon of genes does occasionally result in in-frame mutations which does not lead to knock-down and therefore confounds interpretation of the results. Knock-down efficiency has been shown to vary depending on which domain is targeted, ranging from a strong gene depletion for some domains to minimal phenotype change for others [112]. Interestingly, a screen targeting Cas9 only to exon-encoding domains is an approach to identify drug targets [112]. Screens such as those discussed above will provide researchers with extensive lists of genes involved in the processes of interest, requiring validation. An additional step is the integration of these results with those from other systems-level analyses such as mRNA profiling, proteomics, and phosphoproteomics in order to obtain a more complete picture of the mechanisms underlying cellular processes [108].

2.9.2 CRISPR Interference: CRISPRi

The CRISPRi method was developed to allow targeted disruption of gene expression at the scale of the whole genome [101]. A catalytically dead Cas mutant retains its ability to recognize DNA sequences when in presence of sgRNA but lacks cleavage potential and can interfere with gene transcription via transcriptional elongation, RNA polymerase binding, or transcription factor binding. CRISPRi is thus reversible, in contrast to CRISPRn, which utilizes both Cas' DNA recognition and its cleavage functions and thereby creates a permanent genetic modification. Additionally, a KRAB repression domain can be fused to a deactivated Cas9, thereby repressing gene activity in a targeted manner [113]. A further extension is an inducible version of CRISPRi, namely dCas9-KRAB, which can be used to reversibly inhibit gene expression in a tunable manner [114]. Being in possession of the means to control the expression of nuclease activity is critical to maintaining low but sufficient levels of Cas9 and thereby increasing specificity. Additionally, an inducible nuclease system can also be utilized to precisely control the timing of knock-down, which in the case of human induced pluripotent stem cells has a critical effect on embryonic development and therefore on the outcome of differentiation, and that can be used as readout.

2.9.3 CRISPR Activation: CRISPRa

The CRISPR-Cas system can also be harnessed to perform genome-wide screens with the aim of identifying novel as well as known genes involved in a process of interest. While CRISPRi represses gene expression, clustered regularly interspaced short palindromic repeat activation (CRISPRa) aims to activate gene transcription by targeted delivery of activators to a gene's effector domains [115]. This scalable approach [116] allows modulation of transcription at the level of the genome. The possibility to perform genome-wide loss-of-function and gain-of-function screens with the same setup safe for a different sgRNA collection is a powerful tool, since the effect of gene repression and activation can be compared directly and provide complementary information.

2.9.4 Further Scalable Additions to the CRISPR-Cas Gene Editing Tool Arsenal

One application of the idea of binding an effector to Cas is for the purpose of epigenome editing [117]. In this scenario, a nuclease-dead Cas9 is fused to the catalytic part of human acetyltransferase p300. This allows targeted acetylation of histone H3K27, thereby activating genes via promoters as well as proximal and distal enhancers. While several other enzymes make up the scene of epigenetic control and thus this control is not yet in researchers' hands, the results are promising. The ability to modulate the presence of epigenetic markers would provide researchers with another level of control over gene expression, as well as a precise tool to probe the effect of these dynamic modifications on cellular phenotype.

2.9.5 In vivo Applications

Animal Disease Models

Beyond in vitro cellular manipulation, gene-editing nucleases have large potential in in vivo applications. One immediate impact is the improved ease with which it is possible to generate transgenic models of human diseases in species that were previously difficult to manipulate [59], [60], [118], [119] as well as for engineering human stem cells [33], providing the means to directly assess the consequences of gene disruption and of specific mutations. Reporter and conditional mutant mice were created by injecting Cas9 and different sgRNAs into zygotes in a one-step procedure. Analysis of off-targets in these mice as well as in embryonic stem cells derived from these zygotes concluded that undesired mutations are relatively rare given careful experimental design [120]. Similarly, co-injection of early stage monkey embryos has allowed multiplex genetic engineering, and monkeys harboring two alterations without off-target mutagenesis have been generated [121]. Furthermore, the complexity of the mutational landscape of patients and the interplay with disease could be better understood if the generation of animal models becomes facile to the point of being able to specifically mimic a patient's genetic profile [120], [121].

Gene Therapy

An exciting area of application is the therapeutic field, where targeted nucleases have the potential to reverse disease-related mutations and to disturb or introduce new genes [122], [123]. The potential thereof is immense, proposing to correct genetic as well as non-genetic disorders by reversing mutants, replacing a gene with a corrected version, or knocking down a deleterious gene. Zinc-finger nucleases have been applied to correct mutations responsible for a number of diseases such as sickle cell anemia [124] and others [125], [126]. In addition to the treatment of genetic disorders, preliminary studies have demonstrated the capability of gene-editing nucleases to inactivate genes or introduce protective mutations in the case of non-hereditary diseases, for example, viral infection. One example is the disruption of the CCR5 receptor in lymphocytes by NHEJ, which could be used as a therapy for HIV infection. This and similar approaches have been shown to be possible both with zinc fingers and TALENs [127] as well as CRISPR-Cas9 [128]. A similar strategy has been applied in a clinical trial where the CCR5 gene was knocked out ex vivo by ZF nucleases, with successful outcome [129]. The application of this novel technology in humans as a therapeutic agent will require the lowest possible number of undesired effects and, therefore, an improvement in the current tools for the prediction and experimental measurement of off-target sites. However, current research has provided tools offering a greater potential in limiting unwanted off-target sites, and further specificity could be afforded by coupling this technology with, for example, nanoparticles, thereby localizing genetic interventions to tissues of interest. Moreover, the methods deployed to assess the effects of CRISPR-Cas will need to take into account the complexity of the human body, and calls for understanding the systems and interpretation of the findings.

2.10 A Focus on the Application of Genome-Engineering Nucleases on Chromosomal Rearrangements

One interesting application of CRISPR-Cas technology is the investigation of chromosomal rearrangements from a systems perspective. These rearrangements are the result of joining of two illegitimate partners at the chromosomal level, and are frequently detected in tumors and likely contribute to cancer progression. Given that these rearrangements have been linked to breaks in DNA and faulty cellular repair processes, nucleases are an ideal means to generate DSBs and allow the chromosomal translocations, inversions, or deletions to occur endogenously. Methods developed to assess the extent of Cas9 off-target effects are often also capable of profiling chromosomal rearrangements due to DSBs [79], [87]; large parts of the framework needed for this specific application of nuclease-induced gene editing are already available. In the following sections, we will introduce chromosomal rearrangements and how they are related to disease, and present what has been achieved so far (Fig. 2.2).

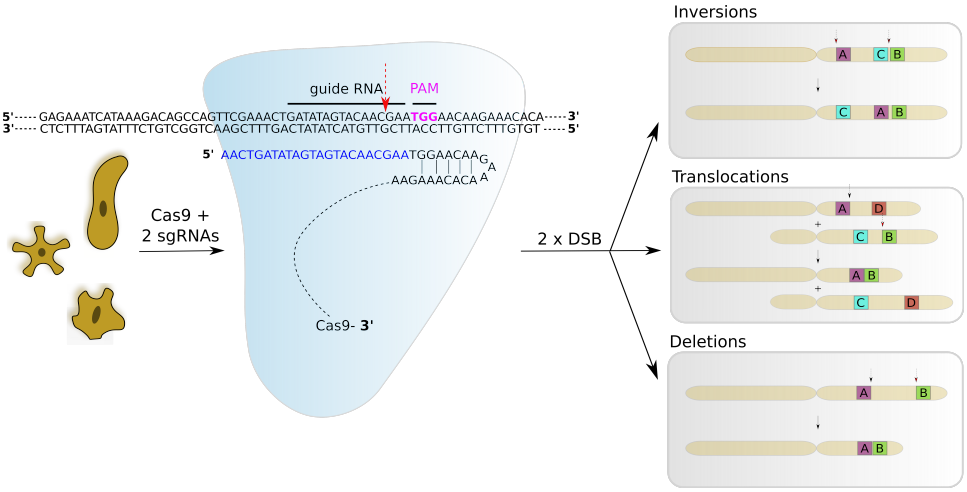


FIGURE 2.2: Different types of chromoso- mal rearrangements can be engineered using nuclease based gene editing tools such as CRISPR-Cas9, e.g. inversions, translocations, and deletions. Briefly, cells expressing the nuclease and two guide RNAs are subject to formation of two DSBs, providing the environment to generate rearrangements. The Cas9 nuclease is targeted to a sequence of interest matching the guide RNA and possessing a 3' PAM site.

2.10.1 Introduction to Chromosomal Rearrangements: The First Disease-Related Translocation

The Philadelphia chromosome was the first chromosomal rearrangement to be consistently associated with a human disease [130]. In the 1980s, researchers discovered that this abnormally small chromosome was caused by an interchange between arms of different chromosomes, characterizing the bcr-abl fusion gene [131]. The t(9;22)(q34;q11) translocation, which produces an oncogenic fusion protein comprised of the N-terminal BCR and C-terminal ABL1, is now known to be the hallmark of chronic myelogenous leukemia (CML). The ABL1 non-receptor tyrosine kinase forms a number of fusion products, which translate into constitutively active tyrosine signaling entities required for cellular transformation [132]. The family of ABL protein kinases associates extracellular stimuli to signaling pathways governing essential cellular mechanisms such as cell growth, migration, DNA damage response, and apoptosis [133]–[135]. Recent studies have demonstrated activation of ABL family members detected in cancer cells derived from melanoma, as well as breast, colon, and lung tissue [136]–[138]. The t(9;22)(q34;q11) chromosomal translocation generates the BCR-ABL1 fusion protein, which leads to oncogenic activation of ABL1 and deregulated tyrosine kinase activity, leading to numerous hematologic malignancies. Several kinase inhibitors have been developed targeting the tyrosine kinase function of ABL1, notably imatinib (Gleevec; Novartis), dasatinib (Sprycel; Bristol-Myers Squibb), and nilotinib (Tasinga; Novartis). Imatinib has been among the more successful treatments of chronic-phase CML and exerts its effect as an ATP competitor. ABL1 has been found to be associated with many other genes forming a variety of fusions, for example, with ETV6, EML1, FOXP1, and others. Generally, the partners with which ABL kinases associate confer the kinase with increased oncogenicity. This also applies to a number of

fusions involving tyrosine kinase moieties, where the fusion partner gene can provide advantageous functionalities, such as targeting the fusion product to the nuclear pore complex, recruiting additional proteins resulting in a larger complex [139], recruiting chaperones mediating protein stability [140], and recruitment of the wild-type form of the partner protein thereby exacerbating the effect of the rearrangement.

In the decades since the original chromosomal translocation involved in leukemia was discovered, myriad translocations have been reported and identified in a number of organs [141]. The unraveling of the BCR-ABL1 story and the successes of therapies based on it pave the way to a better understanding of the arising of chromosomal translocations involving kinase moieties and their impact, as well as to improved drug design in order to target them.

2.10.2 A Global Look at the Mechanisms behind Chromosomal Rearrangements

The detailed mechanism of chromosomal rearrangements is unknown, but the etiology of DNA damage [142] and the ensuing repair of thereby induced DSBs via the NHEJ pathway likely play a large role in the formation of rearrangements. Chromosomal rearrangements are the consequence of the inappropriate religation of two DSBs, caused by endogenous or exogenous sources [143]. Exogenous sources of translocations are ionizing radiation and cancer chemotherapy [144]. A cell can acquire DNA breaks endogenously in a number of ways, for example, after exposure to reactive oxygen species, stalled replication forks leading to free double-strand ends [145], or failure of sister chromatids to separate properly [146]. Alongside these known causes, chromosomal fragile sites (nonrandom loci that exhibit gaps and breaks on metaphase chromosomes following partial inhibition of DNA synthesis) have long been recognized to act as breakage hotspots [147], [148], and have been suggested to be associated with carcinogenic translocations [149], [150]. Both rare and common fragile sites (CFSs) are sensitive to replication stress and are a hotspot for genetic rearrangements, but the most relevant to cancer are the CFSs. Common chromosomal sites are present in all individuals, and the slowing down of replication forks at CFSs as well as a paucity of replication origins [151] suggests that these sites are indeed particularly vulnerable to oncogene-induced replicative stress. Studies have confirmed that a small but significant number of deletions as well as the majority of recurrent cancer-specific fusions are related to CFSs [152]–[155]. A study by Burrow *et al.* [153] determined that, in over half of the gene sets involved in cancer-related translocations, at least one of two partners possesses a breakpoint located at a fragile site. Another bioinformatic study [156] concluded that the observation of breaks occurring preferentially in larger introns is the main significant feature, consistent with the hypothesis that translocations happen after random breakage at fragile sites. However, no specific breakpoint sequence has been discovered, and no sequence or chromatin features were found to be significantly enriched in translocated introns after correcting for intron size.

2.11 A Focus on the Application of Genome-Engineering Nucleases

Interestingly, the genes participating in translocations showed enrichment in "DNA binding" and "transcription-factor binding" ontologies (approximately half of studied translocations compared to 18% in non-translocation-prone genes). Another noteworthy point is the effect of genome organization on translocation frequency. Given that DSBs have limited mobility once formed and that the genome is organized in a nonrandom manner [157], there is evidence for a correlation between spatial proximity of chromosomes and the frequency of translocations [158]–[161]. Cell-type-specific genome organization provides an explanation for the detection of fusions dependent on tissue type. However, studies have indicated that in rare cases partners may find each other over a longer range, suggesting that all translocations may not be formed following the same system [162]. Single-cell tracking experiments have shown that the two ends created by the break migrate together and only dissociate once a partner has been chosen. Once translocation partners find themselves in close proximity, they undergo cycles of pairing and dissociation and those remaining in close proximity are then candidates for NHEJ and the formation of a translocation [162]. As mentioned earlier, the cell possesses various repair mechanisms to handle DSBs, notably homologous recombination, or NHEJ. Recently it has been shown that NHEJ is the predominant DSB repair pathway producing chromosomal translocations, indeed nearly always leaving alterations at the site of DNA repair [163], [164]. Given the strong ties between gene rearrangements and cancer, an improved understanding on the underlying mechanisms of rearrangements and a timeline of the events with respect to disease progression is of great interest, and the involvement of DSBs makes it an ideal application for the CRISPR-Cas platform.

2.11.1 Creating Chromosomal Rearrangements Using CRISPR-Cas

Chromosomal rearrangements can be induced by introducing two DSBs at distinct sites on the same or on different chromosomes [64]–[68], [165], an ideal application for sequence-specific nucleases. Torres and colleagues proved that a CRISPR-Cas setup is suitable for creating chromosomal rearrangements; in separate experiments, they induced two translocation events by directing Cas9 to cleave the genome at two specific loci for each translocation. Choi *et al.* performed similar experiments, showing that not only translocations but also inversions could be created. Although successful, the frequencies with which translocations were detected were low, in accordance with estimates in the occurrence of translocations in vivo [162]. For this reason, both studies introduced a selective marker before tedious screening of single cells to detect the desired fusion event. Shortly thereafter, Maddalo and colleagues successfully induced the EML4–ALK fusion in vivo in mice via adenoviral infection, and observed the appearance of tumors in infected animals. Moreover, known phosphorylation events typical of the deregulation of the ALK oncogene were confirmed, thereby supporting the use of this technique to generate relevant animal disease models. It should be noted that a better

understanding of the extent to which the alternative fusion products interfere with the interpretation of results will be necessary for future studies.

2.12 Future Perspectives

From the genome engineering field's beginnings using recombinant DNA up to the currently popular technique known as CRISPR-Cas9, it is interesting to note that most, if not all, approaches have been heavily inspired directly by already existing biological systems. Continuing basic research will likely uncover more strategies adaptable for the purpose of biological engineering. These same ingenious processes are however also subject to the interconnectedness and the complex rules governing biological systems. In light of this, current and future applications of any nuclease-based technology will require thorough investigation of the effects of its implementation, such as off-target nuclease activity. This is relevant both for academic research projects and for clinical applications, although the impact of the latter is far more apparent. Indeed, off-target genetic alterations not accounted for may lead to deleterious nonreversible changes, given that nucleases operate at the DNA level. Furthermore, the ease with which CRISPR-Cas experiments can be designed makes it an ideal candidate for systems studies, where the whole genome can be probed with a specific perturbation, for example, drug, stimulus, and so on. This also obliges the scientific community to recognize that non-specific side effects must be systematically investigated at a whole-genome level. Currently, the investigation of chromosomal rearrangements and their impact is not yet fully scalable and presents new difficulties such as the effects of the remaining possible gene combinations in addition to the desired rearrangement. Although these have in part been addressed in recent large-scale studies such as loss-of-function screens, improvements in methods for the prediction and detection of nuclease-induced gene alterations will allow increased confidence in the reported results. This is likely to be within reasonable reach in the context of studies such as systems-level gene knockouts. In addition, the improvement of computational methods for the analysis of nonlinear off-target effects will prove critical to deconvolute the results of systems studies and to aid the therapeutic development of this technology.

2.13 References

[1] Capecchi, Mario, "Altering the Genome by Homologous Recombination", *Science*, no. 4910, pp. 1288–1292, 1989.

[2] A. Gossler, T. Doetschman, R. Korn, E. Serflingt, and R. Kemler, "Transgenesis by means of blastocyst-derived embryonic stem cell lines", *Developmental Biology*, vol. 83, no. 23, pp. 9065–9069, 1986.

[3] S. Thompson, A. R. Clarke, A. M. Pow, M. L. Hooper, and D. W. Melton, "Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells", *Cell*, no. 2, pp. 313–21, 1989.

[4] N. Rudin, E. Sugarman, and J. E. Haber, "Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae.", *Genetics*, no. 3, pp. 519–34, 1989.

[5] S. A. Strobel, L. A. Doucette-Stamm, L. Riba, and D. E. Housman, "Site-Specific Cleavage of Human Chromosome 4 Mediated by Triplex-Helix Formation.", *Dervan Source: Science, New Series*, vol. 254, no. 13, pp. 1639–1642, 1991.

[6] J. Cho, M. E. Parks, and P. B. Dervan, "Cyclic polyamides for recognition in the minor groove of DNA.", *Chemistry*, vol. 92, pp. 10 389–10 392, 1995.

[7] A. F. Faruqi, M. M. Seidman, D. J. Segal, D. Carroll, and P. M. Glazer, "Recombination Induced by Triple-Helix-Targeted DNA Damage in Mammalian Cells.", *Molecular and Cellular Biology*, vol. 16, no. 12, pp. 6820–6828, 1996.

[8] J. Gottesfeld, L. Neely, J. Trauger, E. Baird, and P. Dervan, "Regulation of gene expression by small molecules.", *Nature*, vol. 387, no. 6629, pp. 202–205, 1997.

[9] J. Lazowskal, B. Meunier, and C. Macadre, "Homing of a group 11 intron in yeast mitochondrial DNA is accompanied by unidirectional coconversion of upstream-located markers.", *The EMBO Journal*, vol. 13, no. 20, pp. 4963–4972, 1994.

[10] J. V. Moran, S. Zimmerly, R. Eskes, J. C. Kennell, A. M. Lambowitz, R. A. Butow, and P. S. Perlman, "Mobile Group II Introns of Yeast Mitochondrial DNA Are Novel Site-Specific Retroelements.", *Molecular and Cellular Biology*, vol. 15, no. 5, pp. 2828–2838, 1995.

[11] S. Zimmerly, H. Guo, R. Eskest, J. Yang, P. S. Perlman, and A. M. Lambowitz, "A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility.", *Cell*, no. 4, pp. 529–38, 1995.

[12] J. Yang, S. Zimmerly, P. Perlman, and A. Lambowitz, "Efficient integration of an intron RNA into double-stranded DNA by reverse splicing.", *Nature*, no. 6580, pp. 332–5, 1996.

[13] B. Sullenger and T. Cech, "Ribozyme-mediated repair of defective mRNA by targeted trans-splicing.", *Letters to Nature*, no. 6498, pp. 619–22, 1994.

[14] B. S. Chevalier, T. Kortemme, M. S. Chadsey, D. Baker, R. J. Monnat, and B. L. Stoddard, "Design, Activity, and Structure of a Highly Specific Artificial Endonuclease.", *Molecular Cell*, vol. 10, pp. 895–905, 2002.

[15] M. Bibikova, D. Carroll, D. J. Segal, J. K. Trautman, J. Smith, Y.-G. Kim, and S. Chandrasegaran, "Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases.", *Molecular Cell Biology*, vol. 21, no. 1, pp. 289–297, 2001.

[16] J. Miller, A. D. McLachlan, and A. Klug, "Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes.", *Journal of Trace Elements in Experimental Medicine*, no. 6, pp. 1609–14, 1985.

[17] N. Pavletich and C. Pabo, "Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 A.", *Science*, no. 5007, pp. 809–17, 1991.

[18] Y.-G. Kim, J. Cha, and S. Chandrasegaran, "Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain.", *Biochemistry*, vol. 93, no. 3, pp. 1156–1160, 1996.

[19] R. R. Beerli and C. F. Barbas, "Engineering polydactyl zinc-finger transcription factors.", *Nature Biotechnology*, no. 2, pp. 135–41, 2002.

[20] M. Bibikova, M. Golic, K. G. Golic, and D. Carroll, "Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases.", *Genetics*, no. 3, pp. 1169–75, 2002.

[21] J. K. Hye, H. J. Lee, H. Kim, S. W. Cho, and J. S. Kim, "Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly.", *Genome Research*, no. 7, pp. 1279–88, 2009.

[22] F. D. Urnov, E. J. Rebar, M. C. Holmes, H. Steve Zhang, and P. D. Gregory, "Genome editing with engineered zinc finger nucleases.", *Nature Reviews Genetics*, vol. 11, no. 9, pp. 636–46, 2010.

[23] J. Keith Joung and J. D. Sander, "TALENs: a widely applicable technology for targeted genome editing.", *Nature Reviews Molecular Cell Biology*, vol. 14, no. 1, pp. 49–55, 2012.

[24] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, and U. Bonas, "Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors.", *Science*, no. 5959, pp. 1509–12, 2009.

[25] M. J. Moscou and A. J. Bogdanove, "A Simple Cipher Governs DNA Recognition by TAL Effectors.", *Science*, no. 5959, p. 1501, 2009.

[26] K. Miller-Jensen, K. A. Janes, J. S. Brugge, and D. A. Lauffenburger, "Common effector processing mediates cell-specific responses to stimuli.", *Nature*, no. 7153, pp. 604–8, 2007.

[27] C. Mussolino, R. Morbitzer, F. Lütge, N. Dannemann, T. Lahaye, and T. Cathomen, "A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity.", *Nucleic Acids Research*, no. 21, pp. 9283–93, 2011.

[28] T. Cermak, E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove, and D. F. Voytas, "Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting.", *Nucleic Acids Res.*, no. 12, 2011.

[29] J. L. Schmid-Burgk, T. Schmidt, V. Kaiser, and V. Hornung, "A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes.", *Nature Biotechnology*, vol. 31, no. 1, pp. 76–81, 2012.

[30] T. Gaj, C. A. Gersbach, and C. F. Barbas, "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering.", *Trends in Biotechnology*, no. 7, pp. 397–405, 2013.

[31] M. Christian, T. Cermak, E. L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A. J. Bogdanove, and D. F. Voytas, "Targeting DNA double-strand breaks with TAL effector nucleases.", *Genetics*, pp. 757–761, 2010.

[32] J. Guilinger, V. Pattanayak, D. Reyon, S. Tsai, J. Sander, J. Joung, and D. Liu, "Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity.", *Nature Methods*, no. 4, pp. 429–35, 2014.

[33] D. Hockemeyer, H. Wang, S. Kiani, C. S. Lai, Q. Gao, J. P. Cassady, G. J. Cost, L. Zhang, Y. Santiago, J. C. Miller, B. Zeitler, J. M. Cherone, X. Meng, S. J. Hinkley, E. J. Rebar, P. D. Gregory, F. D. Urnov, and R. Jaenisch, "Genetic engineering of human pluripotent cells using TALE nucleases.", *Nature Biotechnology*, no. 8, pp. 731–4, 2011.

[34] Y. Kim, J. Kweon, A. Kim, J. Kyung Chon, J. Yeon Yoo, H. Joo Kim, S. Kim, C. Lee, E. Jeong, E. Chung, D. Kim, M. Seon Lee, E. Mi Go, H. Jung Song, H. Kim, N. Cho, D. Bang, S. Kim, and J.-S. Kim, "A library of TAL effector nucleases spanning the human genome.", *Nature Biotechnology*, vol. 31, no. 3, pp. 251–8, 2013.

[35] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. Romero, and P. Horvath, "CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes.", *Science*, no. 5819, pp. 1709–12, 2007.

[36] R. Barrangou and L. A. Marraffini, "CRISPR-cas systems: Prokaryotes upgrade to adaptive immunity.", *Molecular Cell*, no. 2, pp. 234–44, 2014.

[37] J. van der Oost, E. R. Westra, R. N. Jackson, and B. Wiedenheft, "Unravelling the structural and mechanistic basis of CRISPR Cas systems.", pp. 479–92, 2014.

[38] J. Bondy-Denomy and A. R. Davidson, "To acquire or resist: The complex biological effects of CRISPR-Cas systems.", *Trends in Microbiology*, no. 4, pp. 218–25, 2014.

[39] R. Barrangou and P. Horvath, "CRISPR: New Horizons in Phage Resistance and Strain Identification.", *Annu. Rev. Food Sci. Technol*, vol. 3, pp. 143–62, 2012.

[40] D. Carroll, "Genome Engineering with Targetable Nucleases.", *Annu. Rev. Biochem*, vol. 83, pp. 409–39, 2014.

[41] D. H. Haft, J. Selengut, E. F. Mongodin, and K. E. Nelson, "A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/cas subtypes exist in prokaryotic genomes.", *PLoS Computational Biology*, no. 6, e60, 2005.

[42] K. S. Makarova, D. H. Haft, R. Barrangou, S. J. J Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. M Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost, and E. V. Koonin, "Evolution and classification of the CRISPRCas systems.", *Nature Reviews Microbiology*, pp. 467–77, 2011.

[43] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A Programmable Dual-RNAGuided DNA Endonuclease in Adaptive Bacterial Immunity.", *Science*, no. 6096, pp. 816–21, 2012.

[44] G. Gasiunas, R. Barrangou, P. Horvath, and V. Siksnys, "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.", *Proc Natl Acad Sci U S A.*, no. 39, E2579–86, 2012.

[45] R. Sapranauskas, G. Gasiunas, C. Fremaux, R. Barrangou, P. Horvath, and V. Siksnys, "The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli.", *Nucleic Acids Research*, no. 21, pp. 9275–82, 2011.

[46] H. Deveau, R. Barrangou, J. E. Garneau, J. Labonté, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath, and S. Moineau, "Phage response to CRISPR-encoded resistance in Streptococcus thermophilus.", *Journal of Bacteriology*, no. 4, pp. 1390–1400, 2008.

[47] S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, and J. A. Doudna, "DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.", *Nature*, no. 7490, pp. 62–7, 2014.

[48] M. Jinek, F. Jiang, D. W. Taylor, S. H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, M. Kaplan, A. T. Iavarone, E. Charpentier, E. Nogales, and J. A. Doudna, "Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation.", *Science*, no. 6176, p. 1 247 997, 2014.

[49] H. Nishimasu, F. A. Ran, P. D. Hsu, S. Konermann, S. I. Shehata, N. Dohmae, R. Ishitani, F. Zhang, and O. Nureki, "Crystal structure of Cas9 in complex with guide RNA and target DNA.", *Cell*, no. 5, pp. 935–49, 2014.

[50] L. Cong, A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. Hsu, X. Wu, W. Jiang, L. Marraffini, and F. Zhang, "Multiplex Genome Engineering Using CRISPR/Cas Systems.", *Science*, 2013.

[51] P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. Dicarlo, J. E. Norville, and G. M. Church, "RNA-Guided Human Genome Engineering via Cas9.", *Science*, no. 6121, pp. 823–6, 2013.

[52] P. Mali, K. Esvelt, and G. Church, "Cas9 as a versatile tool for engineering biology.", *Nature Methods*, vol. 10, no. 10, pp. 957–63, 2013.

[53] M. Jinek, A. East, A. Cheng, S. Lin, E. Ma, and J. Doudna, "RNA-programmed genome editing in human cells.", *eLife*, vol. 2, e00471, 2013.

[54] B. R. Adams, A. J. Hawkins, L. F. Povirk, and K. Valerie, "ATM-independent, high-fidelity nonhomologous end joining predominates in human embryonic stem cells.", *Aging*, no. 9, pp. 582–96, 2010.

[55] E. A. Moehle, J. M. Rock, Y.-L. Lee, Y. Jouvenot, R. C. Dekelver, P. D. Gregory, F. D. Urnov, and M. C. Holmes, "Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases.", *Proc Natl Acad Sci U S A.*, no. 9, pp. 3055–60, 2007.

[56] S. Durai, M. Mani, K. Kandavelou, J. Wu, M. H. Porteus, and S. Chandrasegaran, "Zinc finger nucleases: Custom-designed molecular scissors for genome engineering of plant and mammalian cells.", *Nucleic Acids Research*, no. 18, pp. 5978–90, 2005.

[57] C. Öllü, K. Pars, T. I. Cornu, S. Thibodeau-Beganny, M. L. Maeder, J. K. Joung, R. Heilbronn, and T. Cathomen, "Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion.", *Nucleic Acids Research*, no. 22, pp. 8269–76, 2010.

[58] Y. Santiago, E. Chan, P.-Q. Liu, S. Orlando, L. Zhang, F. D. Urnov, M. C. Holmes, D. Guschin, A. Waite, J. C. Miller, E. J. Rebar, P. D. Gregory, A. Klug, and T. N. Collingwood, "Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases.", *Proc Natl Acad Sci U S A.*, no. 15, pp. 5809–14, 2008.

[59] D. F. Carlson, W. Tan, S. G. Lillico, D. Stverakova, C. Proudfoot, M. Christian, D. F. Voytas, C. R. Long, C. Bruce, A. Whitelaw, S. C. Fahrenkrug, and M. Roberts, "Efficient TALEN-mediated gene knockout in livestock.", *Proc Natl Acad Sci U S A.*, no. 43, pp. 17 382–17 387, 2012.

[60] S. Ma, S. Zhang, F. Wang, Y. Liu, Y. Liu, H. Xu, C. Liu, Y. Lin, P. Zhao, and Q. Xia, "Highly Efficient and Specific Genome Editing in Silkworm Using Custom TALENs.", *PLoS ONE*, no. 9, e45035, 2012.

[61] A. Hendel, E. J. Fine, G. Bao, and M. H. Porteus, "Quantifying on-and off-target genome editing.", *Trends in Biotechnology*, no. 2, pp. 132–40, 2015.

[62] H. J. Lee, E. Kim, and J. S. Kim, "Targeted chromosomal deletions in human cells using zinc finger nucleases.", *Genome Research*, no. 1, pp. 81–9, 2010.

[63] M. J. Lee, A. S. Ye, A. K. Gardino, A. M. Heijink, P. K. Sorger, G. MacBeath, and M. B. Yaffe, "Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks.", *Cell*, no. 4, pp. 780–94, 2012.

[64] E. Brunet, D. Simsek, M. Tomishima, R. Dekelver, V. M. Choi, P. Gregory, F. Urnov, D. M. Weinstock, and M. Jasin, "Chromosomal translocations induced at specified loci in human stem cells.", *Proc Natl Acad Sci U S A.*, vol. 106, no. 26, pp. 10 620–5, 2009.

[65] E. H. Breese, C. Buechele, C. Dawson, M. L. Cleary, and M. H. Porteus, "Use of genome engineering to create patient specific MLL translocations in primary human hematopoietic stem and progenitor cells.", *PLoS ONE*, no. 9, e0136644, 2015.

[66] P. S. Choi and M. Meyerson, "Targeted genomic rearrangements using CRISPR/Cas technology.", *Nature Communications*, vol. 5, 2014.

[67] R. Torres, M. C. Martin, A. Garcia, J. C. Cigudosa, J. C. Ramirez, and S. Rodriguez-Perales, "Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPRCas9 system.", *Nature Communications*, vol. 5, no. 3964, 2014.

[68] D. Maddalo, E. Manchado, C. P. Concepcion, C. Bonetti, J. A. Vidigal, Y.-C. Han, P. Ogrodowski, A. Crippa, N. Rekhtman, E. De Stanchina, S. W. Lowe, and A. Ventura, "In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system.", *Nature*, vol. 516, no. 7531, pp. 423–7, 2014.

[69] E. Kim, S. Kim, D. H. Kim, B. S. Choi, I. Y. Choi, and J. S. Kim, "Precision genome engineering with programmable DNA-nicking enzymes.", *Genome Research*, no. 7, pp. 1327–33, 2012.

[70] C. L. Ramirez, M. T. Certo, C. Mussolino, M. J. Goodwin, T. J. Cradick, A. P. McCaffrey, T. Cathomen, A. M. Scharenberg, and J. K. Joung, "Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects.", *Nucleic Acids Research*, no. 12, pp. 5560–8, 2012.

[71] J. Wang, G. Friedman, Y. Doyon, N. S. Wang, C. J. Li, J. C. Miller, K. L. Hua, J. J. Yan, J. E. Babiarz, P. D. Gregory, and M. C. Holmes, "Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme.", *Genome Research*, no. 7, pp. 1316–26, 2012.

[72] H. PD, S. DA, W. JA, R. FA, K. S, A. V, L. Y, F. EJ, W. X, S. O, C. TJ, M. LA, B. G, and Z. F., "DNA targeting specificity of RNA-guided Cas9 nucleases.", *Nat Biotechnol.*, vol. 9, no. 31, pp. 827–32, 2013.

[73] M. J. Lee and M. B. Yaffe, "Protein regulation in signal transduction.", *Cold Spring Harbor Perspectives in Biology*, no. 6, a005918, 2016.

[74] J. G. Doench, E. Hartenian, D. B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B. L. Ebert, R. J. Xavier, and D. E. Root, "Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation.", *Nature Biotechnology*, no. 12, pp. 1262–7, 2014.

[75] R. Chari, P. Mali, M. Moosburner, and G. M. Church, "Unraveling CRISPR-cas9 genome engineering parameters via a library-on-library approach.", *Nature Methods*, no. 9, pp. 823–829, 2015.

[76] D. Y. Guschin, A. J. Waite, G. E. Katibah, J. C. Miller, M. C. Holmes, and E. J. Rebar, "A rapid and general assay for monitoring endogenous gene modification.", *Methods in Molecular Biology*, pp. 247–56, 2010.

[77] J. T. Hill, B. L. Demarest, B. W. Bisgrove, Y. C. Su, M. Smith, and H. J. Yost, "Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products.", *Developmental Dynamics*, no. 12, pp. 1632–6, 2014.

[78] E. K. Brinkman, T. Chen, M. Amendola, and B. Van Steensel, "Easy quantitative assessment of genome editing by sequence trace decomposition.", *Nucleic Acids Research*, no. 22, e168, 2014.

[79] N. Crosetto, A. Mitra, M. Silva, M. Bienko, N. Dojer, Q. Wang, E. Karaca, R. Chiarle, M. Skrzypczak, K. Ginalski, P. Pasero, M. Rowicka, and I. Dikic, "Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing.", *Nature Methods*, no. 4, pp. 361–5, 2013.

[80] T. J. Cradick, P. Qiu, C. M. Lee, E. J. Fine, and G. Bao, "COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites.", 2014.

[81] R. Gabriel, A. Lombardo, A. Arens, J. Miller, P. Genovese, C. Kaeppel, A. Nowrouzi, C. Bartholomae, J. Wang, G. Friedman, M. Holmes, P. Gregory, H. Glimm, M. Schmidt, L. Naldini, and C. von Kalle, "An unbiased genome-wide analysis of zinc-finger nuclease specificity.", *Nature Biotechnology*, no. 9, pp. 816–23, 2011.

[82] A. Lombardo, P. Genovese, C. M. Beausejour, S. Colleoni, Y.-L. Lee, K. A. Kim, D. Ando, F. D. Urnov, C. Galli, P. D. Gregory, M. C. Holmes, and L. Naldini, "Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery.", *Nature Biotechnology*, no. 11, pp. 1298–306, 2007.

[83] Y. Doyon, T. Vo, M. Mendel, S. Greenberg, J. Wang, D. Xia, J. Miller, F. Urnov, P. Gregory, and M. Holmes, "Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures.", *Nat Methods*, no. 1, pp. 74–9, 2011.

[84] Z. Yang, C. Steentoft, C. Hauge, L. Hansen, A. L. Thomsen, F. Niola, M. B. Vester-Christensen, M. Frödin, H. Clausen, H. H. Wandall, and E. P. Bennett, "Fast and sensitive detection of indels induced by precise gene targeting.", *Nucleic Acids Research*, no. 9, e59, 2015.

[85] H. L. Fu, R. R. Valiathan, R. Arkwright, A. Sohail, C. Mihai, M. Kumarasiri, K. V. Mahasenan, S. Mobashery, P. Huang, G. Agarwal, and R. Fridman, "Discoidin domain receptors: Unique receptor tyrosine kinases in collagen-mediated signaling.", *Journal of Biological Chemistry*, no. 11, pp. 7430–7, 2013.

[86] V. Pattanayak, S. Lin, J. P. Guilinger, E. Ma, J. A. Doudna, and D. R. Liu, "High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity.", *Nature Biotechnology*, vol. 31, no. 9, pp. 839–43, 2013.

[87] S. Q. Tsai, Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, and A. J. Iafrate, "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.", *Nature Biotechnology*, no. 2, pp. 1–3, 2014.

[88] I. Fonfara, A. Le Rhun, K. Chylinski, K. S. Makarova, A. L. Lécrivain, J. Bzdrenga, E. V. Koonin, and E. Charpentier, "Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems.", *Nucleic Acids Research*, no. 4, pp. 2577–90, 2014.

[89] Y. Lin, T. J. Cradick, M. T. Brown, H. Deshmukh, P. Ranjan, N. Sarode, B. M. Wile, P. M. Vertino, F. J. Stewart, and G. Bao, "CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences.", *Nucleic Acids Research*, no. 11, pp. 7473–85, 2014.

[90] S. W. Cho, S. Kim, Y. Kim, J. Kweon, H. S. Kim, S. Bae, and J. S. Kim, "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases.", *Genome Research*, 2014.

[91] X. Wu, D. A. Scott, A. J. Kriz, A. C. Chiu, P. D. Hsu, D. B. Dadon, A. W. Cheng, A. E. Trevino, S. Konermann, S. Chen, R. Jaenisch, F. Zhang, and P. A. Sharp, "Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells.", *Nature Biotechnology*, no. 7, pp. 670–6, 2014.

[92] F. A. Ran, P. D. Hsu, C. Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang, and F. Zhang, "Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.", *Cell*, no. 6, pp. 1380–9, 2013.

[93] B. Shen, W. Zhang, J. Zhang, J. Zhou, J. Wang, L. Chen, L. Wang, A. Hodgkins, V. Iyer, X. Huang, and W. C. Skarnes, "Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects.", *Nature Methods*, no. 4, pp. 399–402, 2014.

[94] Y. Fu, J. Sander, D. Reyon, V. Cascio, and K. Joung, "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.", *Nature Biotechnology*, no. 3, pp. 279–284, 2014.

[95] F. Heigwer, G. Kerr, and M. Boutros, "E-CRISP: fast CRISPR target site identification.", *Nature Methods*, vol. 12211, no. 2, pp. 122–3, 2014.

[96] S. Bae, J. Park, and J. S. Kim, "Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases.", *Bioinformatics*, no. 10, pp. 1473–5, 2014.

[97] C. Kuscu, S. Arslan, R. Singh, J. Thorpe, and M. Adli, "Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease.", *Nature Biotechnology*, vol. 32, no. 7, pp. 677–83, 2014.

[98] D. Kim, S. Bae, J. Park, E. Kim, S. Kim, and H. Yu, "Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells.", *Nature Methods*, no. 3, pp. 237–43, 2015.

[99] B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G. W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, and B. Huang, "Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system.", *Cell*, pp. 1479–1491, 2013.

[100] Y. Kim, Z. Paroush, K. Nairz, E. Hafen, G. Jiménez, and S. Y. Shvartsman, "Substrate-dependent control of MAPK phosphorylation in vivo.", *Molecular systems biology*, vol. 7, no. 467, p. 467, Feb. 2011.

[101] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.", *Cell*, no. 5, pp. 1173–83, 2013.

[102] Z. Hou, Y. Zhang, N. E. Propson, S. E. Howden, L.-F. Chu, E. J. Sontheimer, and J. A. Thomson, "Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis.", *Proc Natl Acad Sci U S A.*, no. 39, pp. 15 644–9, 2013.

[103] Esvelt, K and Mali, P and Braff, J and Moosburner, M and Yaung, S and Church, G, "Orthogonal Cas9 proteins for RNA-guided gene regulation and editing.", *Nature Methods*, vol. 10, no. 11, pp. 1116–1121, 2013.

[104] T. Wang, J. J. Wei, D. M. Sabatini, E. S. Lander, T. Wang, J. J. Wei, D. M. Sabatini, and E. S. Lander, "Genetic Screens in Human Cells Using the CRISPR-Cas9 System Terms of Use Genetic screens in human cells using the CRISPR/Cas9 system", *Science*, vol. 343, no. 6166, pp. 80–84, 2013.

[105] O. Shalem, N. Sanjana, E. Hartenian, X. Shi, D. Scott, T. Mikkelsen, D. Heckl, B. Ebert, D. Root, J. Doench, and F. Zhang, "Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells.", *Science*, no. 6166, pp. 84–87, 2014.

[106] H. Koike-Yusa, Y. Li, E. Tan, V.-H. Del Castillo, and K. Yusa, "Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library.", *Nature Biotechnology*, no. 3, pp. 267–73, 2014.

[107] S. Chen, N. E. Sanjana, K. Zheng, O. Shalem, K. Lee, X. Shi, D. A. Scott, J. Song, J. Q. Pan, R. Weissleder, H. Lee, F. Zhang, and P. A. Sharp, "Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis.", *Cell*, pp. 1246–1260, 2015.

[108] O. Parnas, M. Jovanovic, T. M. Eisenhaure, R. H. Herbst, A. Dixit, C. J. Ye, D. Przybylski, R. J. Platt, I. Tirosh, N. E. Sanjana, O. Shalem, R. Satija, R. Raychowdhury, P. Mertins, S. A. Carr, F. Zhang, N. Hacohen, and A. Regev, "A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks.", *Cell*, no. 3, pp. 675–86, 2015.

[109] T. Wang, K. Birsov, N. Hughes, K. Krupczak, Y. Post, J. Wei, E. Lander, and D. Sabatini, "Identification and characterization of essential genes in the human genome.", *Science*, no. 6264, pp. 1096–101, 2015.

[110] T. Hart, M. Chandrashekhar, M. Aregger, Z. Steinhart, K. R. Brown, G. MacLeod, M. Mis, M. Zimmermann, A. Fradet-Turcotte, S. Sun, P. Mero, P. Dirks, S. Sidhu, F. P. Roth, O. S. Rissland, D. Durocher, S. Angers, and J. Moffat, "High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities.", *Cell*, no. 6, pp. 1515–26, 2015.

[111] R. J. Platt, S. Chen, Y. Zhou, M. J. Yim, L. Swiech, H. R. Kempton, J. E. Dahlman, O. Parnas, T. M. Eisenhaure, M. Jovanovic, D. B. Graham, S. Jhunjhunwala, M. Heidenreich, R. J. Xavier, R. Langer, D. G. Anderson, N. Hacohen, A. Regev, G. Feng, P. A. Sharp, and F. Zhang, "CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling.", *Cell*, no. 2, pp. 440–55, 2014.

[112] J. Shi, E. Wang, J. Milazzo, Z. Wang, J. Kinney, and C. Vakoc, "Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains.", *Nat Biotech*, no. 6, pp. 661–7, 2015.

[113] Kearns, Genga, Enuameh, Garber, Wolfre, and Maehr, "Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells.", *Development*, vol. 141, pp. 219–223, 2014.

[114] M. A. Mandegar, N. Huebsch, E. B. Frolov, E. Shin, A. Truong, M. P. Olvera, A. H. Chan, Y. Miyaoka, K. Holmes, C. I. Spencer, L. M. Judge, D. E. Gordon, T. V. Eskildsen, J. E. Villalta, M. A. Horlbeck, L. A. Gilbert, N. J. Krogan, S. P. Sheikh, J. S. Weissman, L. S. Qi, P. L. So, and B. R. Conklin, "CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs.", *Cell Stem Cell*, no. 4, pp. 541–53, 2016.

[115] L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, and L. S. Qi, "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.", *Cell*, no. 2, pp. 442–51, 2013.

[116] L. A. Gilbert, M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann, and J. S. Weissman, "Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.", *Cell*, no. 3, pp. 647–61, 2014.

[117] I. B. Hilton, A. M. D, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy, and C. A. Gersbach, "Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers", *Nature Biotechnology*, no. 5, pp. 510–7, 2015.

[118] A. J. Wood, T.-W. Lo, B. Zeitler, C. S. Pickle, E. J. Ralston, A. H. Lee, R. Amora, J. C. Miller, E. Leung, X. Meng, L. Zhang, E. J. Rebar, P. D. Gregory, F. D. Urnov, and B. J. Meyer, "Targeted Genome Editing Across Species Using ZFNs and TALENs.", *Science*, vol. 333, p. 307, 2011.

[119] P. Huang, A. Xiao, M. Zhou, Z. Zhu, S. Lin, and B. Zhang, "Heritable gene targeting in zebrafish using customized TALENs.", *Nature Biotechnology*, no. 8, pp. 699–700, 2011.

[120] H. Wang, H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang, and R. Jaenisch, "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering.", *Cell*, no. 4, pp. 910–8, 2013.

[121] Y. Niu, B. Shen, Y. Cui, Y. Chen, J. Wang, L. Wang, Y. Kang, X. Zhao, W. Si, W. Li, A. P. Xiang, J. Zhou, X. Guo, Y. Bi, C. Si, B. Hu, G. Dong, H. Wang, Z. Zhou, T. Li, T. Tan, X. Pu, F. Wang, S. Ji, Q. Zhou, X. Huang, W. Ji, and J. Sha, "Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos.", *Cell*, no. 4, pp. 836–43, 2014.

[122] Y. Wu, D. Liang, Y. Wang, M. Bai, W. Tang, S. Bao, Z. Yan, D. Li, and J. Li, "Correction of a genetic disease in mouse via use of CRISPR-Cas9.", *Cell Stem Cell*, no. 6, pp. 659–62, 2013.

[123] G. Schwank, B. K. Koo, V. Sasselli, J. F. Dekkers, I. Heo, T. Demircan, N. Sasaki, S. Boymans, E. Cuppen, C. K. Van Der Ent, E. E. S. Nieuwenhuis, J. M. Beekman, and H. Clevers, "Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients.", *Cell Stem Cell*, no. 6, pp. 653–8, 2013.

[124] N. Sun, J. Liang, Z. Abil, and H. Zhao, "Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease.", *Mol. BioSyst. Mol. BioSyst*, vol. 8, no. 8, pp. 1255–1263, 2012.

[125] K. Yusa, S. T. Rashid, H. Strick-Marchand, I. Varela, P.-Q. Liu, D. E. Paschon, E. Miranda, A. Ordóñez, N. R. F. Hannan, F. J. Rouhani, S. Darche, G. Alexander, S. J. Marciniak, N. Fusaki, M. Hasegawa, M. C. Holmes, J. P. Di Santo, D. A. Lomas, A. Bradley, and L. Vallier, "Targeted gene correction of a 1 -antitrypsin deficiency in induced pluripotent stem cells.", *Nature*, no. 7369, pp. 391–4, 2011.

[126] F. Soldner, J. Lagani?re, A. W. Cheng, D. Hockemeyer, Q. Gao, R. Alagappan, V. Khurana, L. I. Golbe, R. H. Myers, S. Lindquist, L. Zhang, D. Guschin, L. K. Fong, B. J. Vu, X. Meng, F. D. Urnov, E. J. Rebar, P. D. Gregory, H. S. Zhang, and R. Jaenisch, "Generation of isogenic pluripotent stem cells differing exclusively at two early onset parkinson point mutations.", *Cell*, no. 2, pp. 318–31, 2011. arXiv: 0402594v3 [cond-mat].

[127] E. E. Perez, J. Wang, J. C. Miller, Y. Jouvenot, K. A. Kim, O. Liu, N. Wang, G. Lee, V. V. Bartsevich, Y.-L. Lee, D. Y. Guschin, I. Rupniewski, A. J. Waite, C. Carpenito, R. G. Carroll, J. S. Orange, F. D. Urnov, E. J. Rebar, D. Ando, P. D. Gregory, J. L. Riley, M. C. Holmes, and C. H. June, "Establishment of HIV-1 resistance in CD4 + T cells by genome editing using zinc-finger nucleases.", *Nat Biotech*, no. 7, pp. 808–16, 2008.

[128] H. Ebina, N. Misawa, Y. Kanemura, and Y. Koyanagi, "Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus.", *Scientific Reports*, 2013.

[129] P. Tebas, D. Stein, W. W. Tang, I. Frank, S. Q. Wang, G. Lee, S. K. Spratt, R. T. Surosky, M. A. Giedlin, G. Nichol, M. C. Holmes, P. D. Gregory, D. G. Ando, M. Kalos, R. G. Collman, G. Binder-Scholl, G. Plesa, W.-T. Hwang, B. L. Levine, and C. H. June, "Gene Editing of CCR5 in Autologous CD4 T Cells of Persons Infected with HIV.", *New England Journal of Medicine*, vol. 37010370, no. 10, pp. 901–10, 2014.

[130] Janet D Rowley, "A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine fluorescence and Giemsa staining.", *Nature*, vol. 243, no. 5404, pp. 290–3, 1973.

[131] A. de Klein, A. G. van Kessel, G. Grosveld, C. R. Bartram, A. Hagemeijer, D. Bootsma, N. K. Spurr, N. Heisterkamp, J. Groffen, and J. R. Stephenson, "A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia.", *Nature*, vol. 300, no. 5894, pp. 765–7, Dec. 1982.

[132] R. Ren, "Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia.", *Nature Reviews*, 2005.

[133] J. Colicelli, "ABL tyrosine kinases: evolution of function, regulation, and specificity.", *Science Signaling*, vol. 3, 2010.

[134] W. D. Bradley and A. J. Koleske, "Regulation of cell migration and morphogenesis by Abl-family kinases: emerging mechanisms and physiological contexts.", *Journal of cell science*, vol. 122, no. Pt 19, pp. 3441–54, Oct. 2009.

[135] A. B. Raitano, Y. E. Whang, and C. L. Sawyers, "Signal transduction by wild-type and leukemogenic Abl proteins.", *Biochimica et Biophysica Acta - Reviews on Cancer*, no. 3, F201–16, 1997.

[136] D. Srinivasan and R. Plattner, "Activation of Abl tyrosine kinases promotes invasion of aggressive breast cancer cells.", *Cancer Research*, no. 11, pp. 5648–55, 2006.

[137] K. Rikova, A. Guo, Q. Zeng, A. Possemato, J. Yu, H. Haack, J. Nardone, K. Lee, C. Reeves, Y. Li, Y. Hu, Z. Tan, M. Stokes, L. Sullivan, J. Mitchell, R. Wetzel, J. MacNeill, J. M. Ren, J. Yuan, C. E. Bakalarski, J. Villen, J. M. Kornhauser, B. Smith, D. Li, X. Zhou, S. P. Gygi, T.-L. L. Gu, R. D. Polakiewicz, J. Rush, and M. J. Comb, "Global Survey of Phosphotyrosine Signaling Identifies Oncogenic Kinases in Lung Cancer", *Cell*, vol. 131, no. 6, pp. 1190–203, 2007.

[138] J. M. Drake, N. A. Graham, T. Stoyanova, A. Sedghi, A. S. Goldstein, H. Cai, D. A. Smith, H. Zhang, E. Komisopoulou, J. Huang, T. G. Graeber, and O. N. Witte, "Oncogene-specific activation of tyrosine kinase networks during prostate cancer progression.", *Proc Natl Acad Sci U S A.*, vol. 109, no. 5, pp. 1643–8, 2012.

[139] K. De Keersmaecker, M. Versele, J. Cools, G. Superti-Furga, and O. Hantschel, "Intrinsic differences between the catalytic properties of the oncogenic NUP214-ABL1 and BCR-ABL1 fusion protein kinases.", *Leukemia*, vol. 22, no. 12, pp. 2208–16, 2008.

[140] C. S. Kasyapa, P. Kunapuli, and J. K. Cowell, "HSPA1A is an important regulator of the stability and function of ZNF198 and its oncogenic derivative, ZNF198-FGFR1.", *Journal of Cellular Biochemistry*, no. 5, pp. 1308–17, 2007.

[141] N. Stransky, E. Cerami, S. Schalm, J. L. Kim, and C. Lengauer, "The landscape of kinase fusions in cancer.", *Nature Communications*, no. 4846, 2014.

[142] A. G. Tsai and M. R. Lieber, "Mechanisms of chromosomal rearrangement in the human genome.", *BMC Genomics*, vol. 11, no. 1, 2010.

[143] M. Byrne, J. Wray, B. Reinert, Y. Wu, J. Nickoloff, S. H. Lee, R. Hromas, and E. Williamson, "Mechanisms of oncogenic chromosomal translocations.", *Annals of the New York Academy of Sciences*, pp. 89–97, 2014.

[144] P. D. Aplan, "Causes of oncogenic chromosomal translocation.", *Trends in Genetics*, no. 1, pp. 46–55, 2006.

[145] C. Allen, A. K. Ashley, R. Hromas, and J. A. Nickoloff, "More forks on the road to replication stress recovery.", *Journal of Molecular Cell Biology*, no. 1, pp. 4–12, 2011.

[146] W. K. Kaufmann, "Dangerous entanglements", *Trends in Molecular Medicine*, no. 6, pp. 235–7, 2006.

[147] H. Gaillard, T. García-Muse, and A. Aguilera, "Replication stress and cancer.", *Nature Reviews Cancer*, vol. 15, no. 5, pp. 276–89, 2015.

[148] S. G. Durkin and T. W. Glover, "Chromosome Fragile Sites.", *Annual Review of Genetics*, pp. 169–92, 2007.

[149] F. Hecht and G. R. Sutherland, "Fragile Sites and Cancer Breakpoints.", *Cancer Genetics and Cytogenetics*, no. 2, pp. 179–81, 1984.

[150] M. M. LeBeau and J. D. Rowley, "Heritable fragile sites in cancer.", *Nature*, no. 5960, pp. 607–8, 1984.

[151] T. W. Glover and T. E. Wilson, "Breaking News on Fragile Sites in Cancer.", *Cancer Cell*, no. 2, pp. 137–9, 2013.

[152] G. R. Bignell, C. D. Greenman, H. Davies, A. P. Butler, S. Edkins, J. M. Andrews, G. Buck, L. Chen, D. Beare, C. Latimer, S. Widaa, J. Hinton, C. Fahey, B. Fu, S. Swamy, G. L. Dalgliesh, B. T. Teh, P. Deloukas, F. Yang, P. J. Campbell, P. A. Futreal, and M. R. Stratton, "Signatures of mutation and selection in the cancer genome.", *Nature*, no. 7283, 2010.

[153] A. A. Burrow, L. E. Williams, L. C. Pierce, and Y.-H. Wang, "Over half of breakpoints in gene pairs involved in cancer-specific recurrent translocations are mapped to human chromosomal fragile sites.", *BMC Genomics*, vol. 10, no. 59, 2009.

[154] R. I. Richards, "Fragile and unstable chromosomes in cancer: causes and consequences.", *Trends in Genetics*, vol. 17, no. 6, pp. 339–45, 2001.

[155] J. M. Fang, M. F. Arlt, A. C. Burgess, S. L. Dagenais, D. G. Beer, and T. W. Glover, "Translocation Breakpoints in FHIT and FRA3B in Both Homologs of Chromosome 3 in an Esophageal Adenocarcinoma.", *Genes Chromosomes and Cancer*, vol. 30, no. 3, pp. 292–8, 2001.

[156] F. J. Novo, "Chromosome translocations in cancer: computational evidence for the random generation of double-strand breaks.", *Trends in Genetics*, vol. 22, no. 4, pp. 193–6, 2006.

[157] T. Misteli, "Beyond the Sequence: Cellular Organization of Genome Function.", *Cell*, vol. 128, pp. 787–800, 2007.

[158] H. Neves, C. Ramos, M. Gomes, D. Silva, A. N. Parreira, and L. Parreira, "The Nuclear Topography of ABL, BCR, PML, and RARa Genes: Evidence for Gene Proximity in Specific Phases of the Cell Cycle and Stages of Hematopoietic Differentiation.", *Blood*, vol. 93, no. 4, pp. 1197–207, 1999.

[159] J. Roix, P. McQueen, P. Munson, L. Parada, and T. Misteli, "Spatial proximity of translocation-prone gene loci in human lymphomas.", *Nature Genetics*, no. 3, pp. 287–91, 2003.

[160] W. A. Bickmore and P. Teague, "Influences of chromosome size, gene density and nuclear position on the frequency of constitutional translocations in the human population.", *Chromosome Research*, no. 8, pp. 707–15, 2002.

[161] Nikiforova, "Proximity of Chromosomal Loci That Participate in Radiation-Induced Rearrangements in Human Cells.", *Science*, vol. 290, no. 5489, pp. 138–41, 2000.

[162] V. Roukos, T. C. Voss, C. K. Schmidt, S. Lee, D. Wangsa, and T. Misteli, "Spatial Dynamics of Chromosome Translocations in Living Cells.", *Science*, vol. 341, no. 6146, pp. 660–4, 2013.

[163] J. A. Nickoloff, L. P. De Haro, J. Wray, and R. Hromas, "Mechanisms of leukemia translocations.", *Curr Opin Hematol .*, vol. 15, no. 4, pp. 338–45, 2008.

[164] J. H. Wang, "Mechanisms and impacts of chromosomal translocations in cancers.", *Frontiers of Medicine*, no. 3, pp. 263–74, 2012.

[165] D. Simsek, E. Brunet, S. Y. W. Wong, S. Katyal, Y. Gao, P. McKinnon, J. Lou, L. Zhang, J. Li, E. Rebar, P. Gregory, M. Holmes, and M. Jasin, "DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation.", *PLoS Genetics*, no. 6, e1002080, 2011.

Part II

Cellular perturbations

Chapter 3

Oncogenic transformation and altered signaling networks upon expression of the PRKAR1A-RET chromosomal translocation in immortalized thyroid epithelial cells

Franziska Voellmy[1,2,3], Craig Simpson[1], Conor Howard[1,4],Gaye Saginc[1,3], Xavier Robin[1,3], James Longden[1,3], Rune Linding[1,3]*

[1] Biotech Research and Innovation Centre, University of Copenhagen, Copenhagen, 2200, Denmark

[2] Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark

[3] Natural History Museum of Denmark, University of Copenhagen, Copenhagen, 1350, Denmark

[4] Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, 94158, United States

* correspondence to linding@lindinglab.org

3.1 Abstract

Chromosomal rearrangements are massive genetic events detected across many blood and solid cancers, both in cultured cells as well as in patients. While some of their interactors or downstream targets have been investigated, little is known at a systems-level about how they drive cells toward a cancerous phenotype. Here we used the CRISPR-Cas9 technology to induce double-strand breaks at two chromosomal locations, thereby creating the conditions necessary for rearrangement. We applied this to generate and endogenously express the thyroid-specific PRKAR1A-RET fusion in immortalized Nthy-ori 3-1 thyroid follicular epithelial cells, increasing the transformative potential of these cells. Mass spectrometry-based analysis found

that the regulated proteins and phosphosites following fusion induction enriched for cancer-related processes, notably regulation of cell cycle control, MYC and mTOR signaling, EMT and hypoxia. In addition to the processes identified prior to transformation, the transformed cells also enriched for apical junctions. These data thus elucidate potential mechanisms by which the PRKAR1A-RET translocation pushed thyroid cells toward a malignant phenotype.

3.2 Introduction

Since the discovery of the first chromosomal translocation BCR-ABL1 and the subsequent success of the ABL tyrosine kinase inhibitor imatinib for the treatment of chronic myelogenous leukemia, chromosomal rearrangements involving kinases have been identified across a broad range of blood and solid tumors. Several of these have been determined to be oncogenic and identified as driving specific processes related to cancer [1]. Despite their potential as biomarkers and drug targets, our understanding of the changes they incur at the level of protein and signaling networks is insufficient to understand their global impact on cancer cells.

In most kinase fusions retaining functional activity, the kinase moiety is located C-terminally with respect to the fusion partner. The N-term fusion partner gene provides the alteration in expression, oligomerization or localization of kinase activity required for oncogenic activity [2]. The frequency of fusions is tissue-dependent, with thyroid cancers displaying the highest percentage of recurring kinase rearrangements. Analyses have shown that fusions with recurrent kinase genes in thyroid cancers are mutually exclusive with respect to other rearrangements and are thus expected to be key events in cancer initiation and progression [1].

A large fraction of the research involving gene fusions has been centered around blood-related malignancies. Moreover, the majority of studies of genetic rearrangements at a larger scale are performed on cells easily amenable to manipulation and results thus still require validation in a more physiologically relevant context [3], [4].

Here we examine the effects of the endogenous expression of the thyroid-specific PRKAR1A-RET translocation in an immortalized thyroid cell line by investigation of the proteome and phosphoproteome by mass spectrometry.

Proto-oncogene tyrosine-protein kinase receptor Ret (RET) is the most common kinase fusion moiety in human thyroid cancer, and associates with PRKAR1A, CCDC6, AKAP13, FKBP15, and NTRK, among others. In the case of the PRKAR1A-RET fusion (Fig. 3.1, panel a), the genomic breakpoint occurs in the intron of the C-terminal partner immediately preceding the exon coding for the tyrosine kinase domain. Two fusion products can occur due to alternative splicing of RET: one 52 aa long, the other 9 aa long (RET portion). The cDNA breakpoint has however been observed to be identical across many patient tissues, preserving the catalytic activity of the tyrosine kinase [5].

The spatial organization of chromosomes varies depending on the tissue of origin, as does the landscape of genetic rearrangements. Indeed, translocations are preferentially formed when smaller distances separate the two fusion partners chromosomes [6]. Despite the etiology behind the genesis of these

aberrations not being fully understood, genetic instability likely plays a role, whether setting the stage for the fusion to occur, or as a result thereof which would then influence further cellular events leading to disease progression. These altered dynamics also depend on the specific cellular context, and we thus introduce the thyroid-specific PRKAR1A-RET fusion in immortalized thyroid epithelial cells, e.g. rather than in traditionally used NIH3T3 cells [7].

We used CRISPR-Cas9 technology to create two double-stranded breaks within the introns of the partner genes reported to harbor the genomic fusion breakpoint [8]–[10]. Following cleavage, rearrangements occur at low frequency [9]. Importantly, this method for generation of chromosomal rearrangement allows for an endogenous expression level of the resulting fusion product, which we expect to lead to a cellular system displaying protein and signaling dynamics with increased resemblance to a physiological setting (Fig. 3.1, panel c), essential for the interpretation of events leading to a cancerous phenotype.

In addition to improving our understanding of how kinase rearrangements alter the proteomic and signaling state of cells, we address whether the endogenous expression of the fusion product is capable of triggering the transformative process. We assess anchorage-independent growth using a soft agar assay, as well as used this as a means for selection of subclones to determine the changes required for malignant transformation. Proliferation in the absence of anchorage has been shown to correlate with tumorigenicity [11]. These colonies were cultured briefly thereafter and cells harvested in order to investigate the altered cell state required for transformation.

3.3 Results

3.3.1 Induction of the PRKAR1A-RET chromosomal rearrangement

To identify and quantify the protein concentration and signaling changes induced upon expression of the thyroid-specific PRKAR1A-RET translocation, we used a label-free mass spectrometry approach. Because normal human epithelial cells senesce after only a few divisions in culture we used SV-40 immortalized non-transformed thyroid cells (Nthy-ori 3-1 [12]) that are suitable for cell culture, but have low colony formation ability and are non-tumorigenic in mice.

The fusion of intron 7 of PRKAR1A (exons 1 through 7) and intron 11 of RET (exons 12 through 20) leads to a fusion product of 596 amino acids in length, retaining the regulatory and cyclic NMP binding domains of the N-terminal PRKAR1A and the full tyrosine kinase domain belonging to the C-terminal RET kinase (Fig. 3.1, panel b). CRISPR-Cas technology was used to generate two double-strand breaks at the intronic locations corresponding to the identified breakpoint thereby allowing the chromosomal rearrangement to occur. Cas9 and the selected guide RNAs (HA, HB, TA, TB) were virally transduced in thyroid cells (Fig. 3.2), and Cas9 cleavage activity was induced (Fig. 3.1, panel c). Endogenous DNA repair led to the formation of fusions at low frequency, thereby generating a mixed population of cells. From these mixed populations we selected cells containing the fusion by single cell cloning. We screened clones for the fusion by PCR using cDNA primers

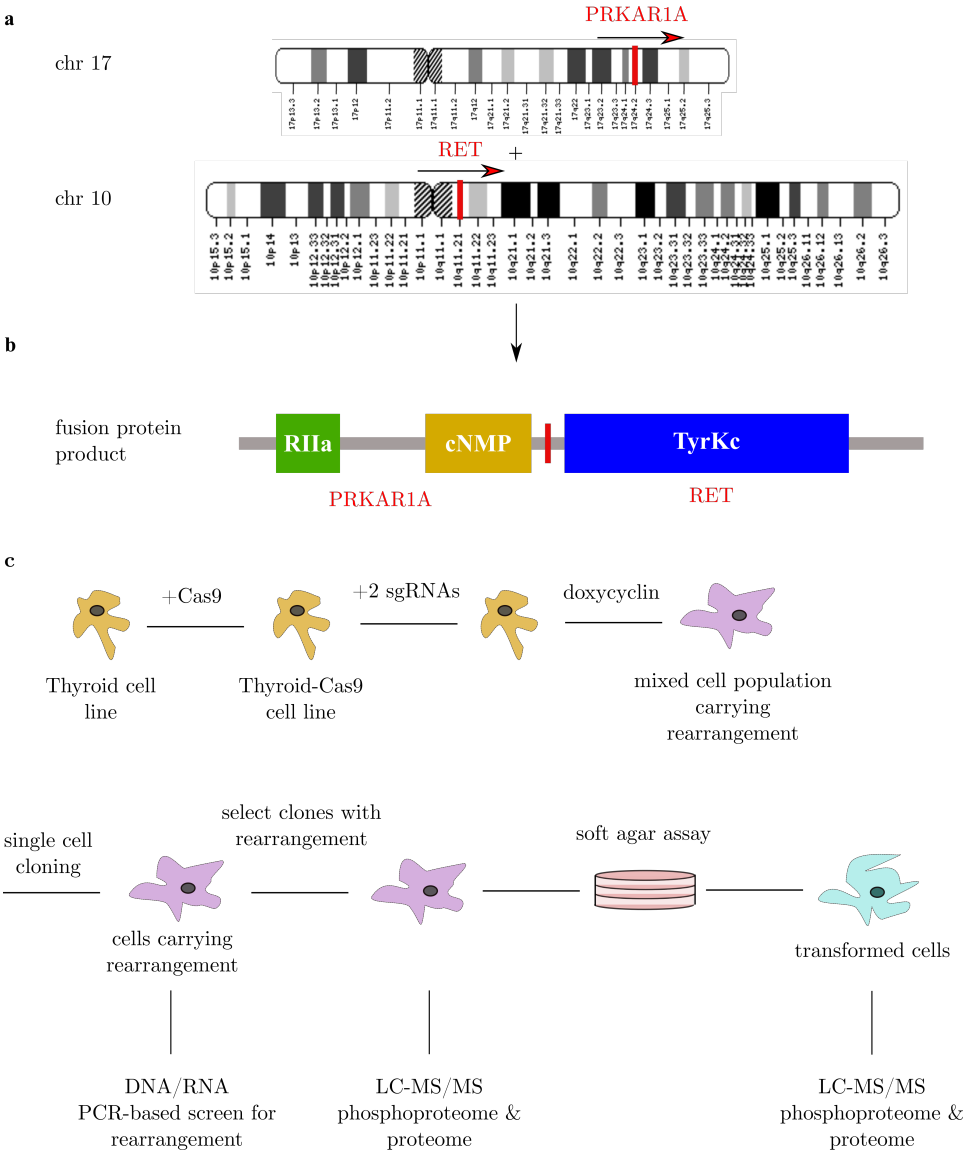


FIGURE 3.1: **a** The translocation formed by the fusion of PRKAR1A and RET genes. The chromosomal breakpoint (shown in red) was situated in introns of chromosomes 17 and 10, respectively. **b** SMART domain prediction for the protein product of the fusion described in **a**. The Regulatory subunit portion of type II PKA R-subunit (RIIa) and the cyclic nucleotide-monophosphate binding domain (cNMP) domains of PRKAR1A were fused N-terminally with the intact tyrosine kinase domain of RET. **c** depicts the main steps in detail leading to the generation of cells used for subsequent mass spectrometric analyses.

surrounding the breakpoint. Due to the requirement of a PAM sequence for directed Cas9 cleavage, the intronic breakpoint did not exactly correspond to the breakpoint reported in tissue samples. However, the resulting cDNA and therefore expressed protein sequence was identical to the sequence reported previously. Four combinations of guide RNAs were transduced targeting the intronic breakpoint region (HATA, HATB, HBTA, HBTB), two of which (HATB, HBTB) generated fusion-positive clones. Cell populations derived from three single cell clones corresponding to each of the two successful guide RNA combinations were chosen for further experiments, resulting in a total of 6 fusion-expressing clones, named here HATB_1, HATB_2, HATB_3, HBTB_1, HBTB_2, HBTB_3. After identification of successful fusion expression, protein was harvested for mass spectrometry-based sampling of the rewired network upon expression of the rearrangement product.

After normalization of peptide intensities based on linear regression, replicates were combined and peptide ratios to Cas9-transduced cells (ThyCas9) were analyzed using a Markov Chain Monte Carlo framework to determine protein concentration and phosphosite occupancy ratios along with a measure of uncertainty. From this analysis we identified 5'456 quantifiable proteins, and 9'723 unique phosphorylated sites, mapping to 2'742 proteins carrying a phosphosite. We subjected proteins and sites quantified over all samples to hierarchical clustering using a Euclidian distance metric shown in in Fig. 3.2. For both the identified proteins and phospopeptides we observed distinct clustering of the fusion-expressing cells compared to the control cells. The remarkably similar signal observed in each fusion-expressing cell line would suggest that chromosomal rearrangements drive quantifiable yet conserved changes at the cell signaling level.

3.3.2 Protein expression and signaling in cells expressing PRKAR1A-RET

We found 1'208 of the quantified proteins (22.1%) to be significantly regulated (z-score 1.2) in cells containing the PRKAR1A-RET fusion compared to the control (z-score 1.2). Overall, we observed a shift toward differential downregulation (797 proteins) compared to upregulation (448 proteins).

We compared our regulated proteins to proteins found to be important in thyroid cancer [13]. Within this set we identified TPR (previously observed to fuse with MET in thyroid cancer), KTN1 (highly expressed in thyroid cancer compared to other cancer types), TPM3 (a tumor suppressor gene), ERC1 (related to the NF-κ B pathway), and NRAS (frequently mutated oncogene). Interestingly several of these (TPR, KTN1, TPM3, ERC1) have been detected as rearrangement partners in thyroid cancer. This is in line with the observation that thyroid cancer displays a high number of fusions in comparison to other solid tumor cancers.

3.3.3 Predicted PPI

We next wished to address whether our network of identified and regulated proteins faithfully replicated the effect of the fusion in thyroid cancer. Given the dearth of systematic studies of gene fusions, the downstream changes as

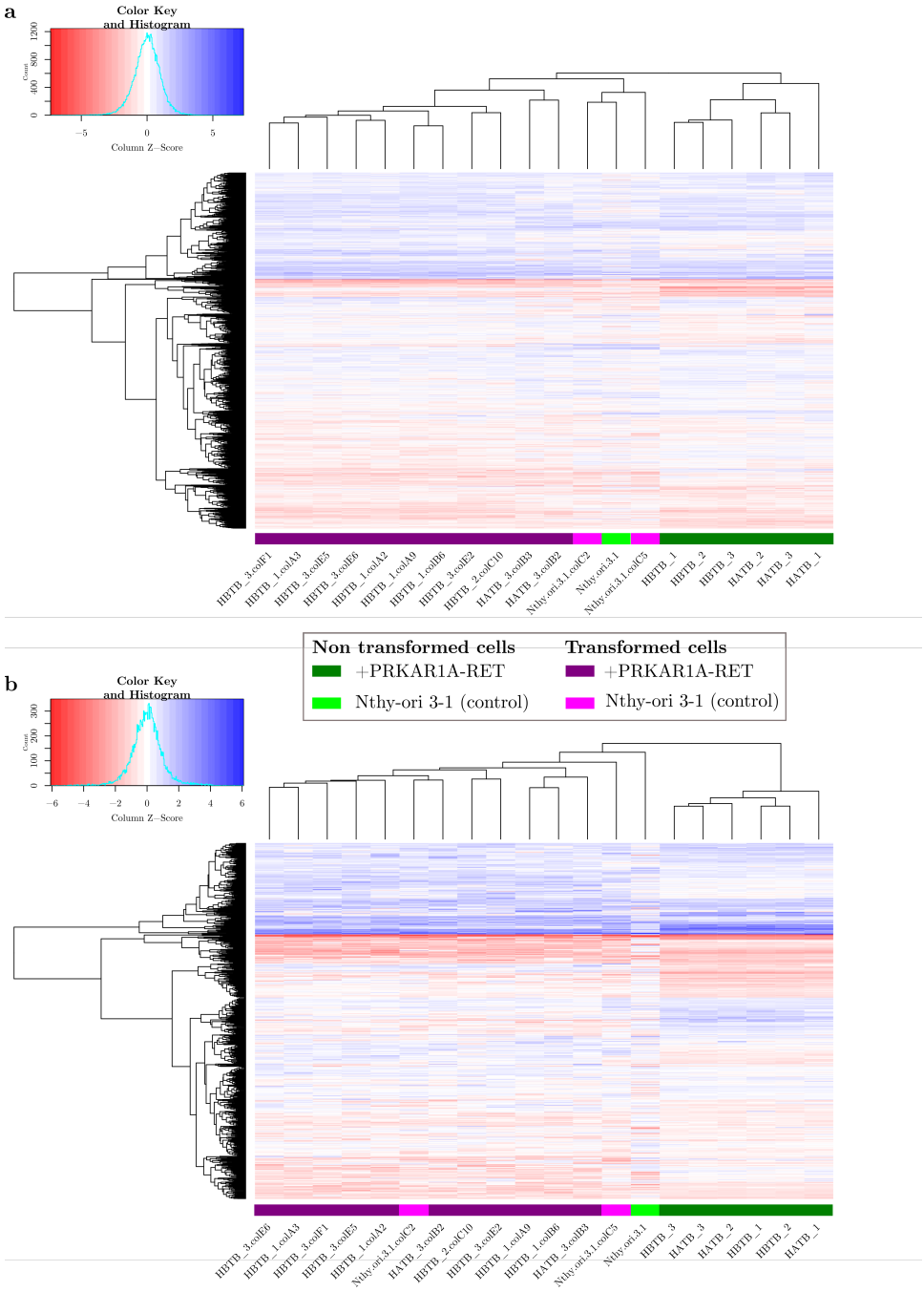


FIGURE 3.2: **a** shows the protein concentration log ratio to ThyCas9 cells for all cells expressing the fusion as well as the original Nthy-ori 3-1 cells. **b** shows phospho-site occupancy log ratios for the same sample set as in **a**. Only proteins or sites seen in all samples are represented here.

Gene name	Ensembl Protein ID
NPM1	ENSP00000428755
HSP90AA1 (**)	ENSP00000335153
HNRNPD	ENSP00000313199
PDLIM7	ENSP00000348099
EGFR	ENSP00000275493
STOML2	ENSP00000348886
PPP1CA	ENSP00000326031
PRKACA	ENSP00000309591
AIP	ENSP00000279146
NELL2	ENSP00000416341
CDK2	ENSP00000452514
VAPA (**)	ENSP00000345656
CRK	ENSP00000300574
PLCG1	ENSP00000244007
IKBKG	ENSP00000483825

TABLE 3.1: Predicted interactors of PRKAR1A-RET (ChiPPI) found to be significantly regulated upon fusion expression. **Bold font** denotes the tumor suppressor genes within the overlap. *Italic* denote known oncogenes. (**) denotes proteins with phosphosite regulation.

well as the interactors of the PRKAR1A-RET fusion are not well characterized. The ChiPPi database addresses this issue by predicting the protein-protein interactions of a fusion based on known preferential interaction of protein domains and their co-occurrence and thus predicts the resulting affected proteins [14]. Following this strategy, predictions suggest interactions may be gained (oncogenes) or lost (tumor suppressors) upon fusion expression, leading to a more cancerous cellular state. ChiPPI predicted 111 unique interactions of the parental genes (here PRKAR1A and RET) to be affected by the generation of the fusion. Of these, we identify 70 (63%) in our dataset, 73% of which are also identified the phosphorylation data. Furthermore, of the predicted proteins identified in the concentration data 15 showed differential expression in our dataset (z-score 1.2) when comparing changes in the cells containing the fusion compared to the ThyCas9 cells (Table 3.1). Of the 15 regulated proteins that were also predicted to be affected, 4 (26.6%) were tumor-suppressor genes (ChiPPI predicted 20 TSGs to be modulated). We identified 5 out of the 8 protein-protein interactions predicted to be lost upon expression of the fusion, with none of these being significantly regulated in cells containing the fusion. Two of the predicted altered interactors also showed significant phosphosite regulation. The overall agreement of our data with that of predicted changes due to this fusion consolidated our confidence in the dataset.

3.3.4 Fusion expression leads to enriched hallmarks

In order to determine whether the observed signaling network changes driven by PRKAR1A-RET fusion were related to oncogenesis we assigned functional annotations using the hallmark gene set collection of MSigDB, based on coherently expressed genes representative of biological processes [15]. We compared the regulated proteins and phosphoproteins in cells carrying the fusion to ThyCas9 cells in terms of their hallmarks. We found that proteins regulated in cells harboring the fusion were extremely highly enriched for MYC targets (FDR = 3.37x10-111) (Fig. 3.3). The overexpression and amplification of MYC is involved in the majority of human cancers and leads to the acquisition of cancer hallmark characteristics such as uncontrolled proliferation. We also observed a high enrichment for proteins upregulated by mTOR related signaling (FDR = 3.06x10-71), which have been shown to affect cell proliferation, viability as well as iodide uptake in thyroid cells [16]. Several other processes central to cancer were found to be strongly regulated in our dataset: oxidative phosphorylation (FDR = 3.03x10-42), targets of E2F transcription factors (FDR = 1.85x10-35), genes involved in the G2/M cell cycle checkpoint (FDR = 3.96x10-32), epithelial-mesenchymal transition (FDR = 8.57x10-24), as well as hypoxia (FDR = 6.88x10-21). Enrichment for genes regulated by E2F transcription factors as well as the G2/M cell cycle checkpoint are both indicative of changes occurring at the level of the cell cycle, crucial to the cancer cell's ability to sustain uncontrolled growth. Epithelial-mesenchymal transition is a major mechanism involved in tumor progression and has been reported to be activated in anaplastic thryoid carcinomas [17].

Perhaps surprisingly we identified only 54 significantly modulated phosphoproteins following expression of the PRKAR1A-RET fusion, which could be

linked to the nature of the experiment. Indeed, after the formation of the rearrangement following nuclease-mediated double-strand breakage, the fusion product was under the control of the endogenous promoter of the 5' translocation partner. This is in contrast to studies where the fusion construct is constantly (over)expressed and therefore unsurprisingly elicits high levels of signaling. We nevertheless sought to investigate the nature of the regulated phosphosites.

Regulated phosphosites showed an enrichment for similar processes as differentially regulated proteins with MYC targets (FDR = 1.22x10-7), E2F targets (p-value = 8.63x10-5), the G2/M checkpoint (FDR = 8.63x10-5), protein secretion (FDR = 1.92x10-4) and epithelial-mesenchymal transition (FDR = 1.62x10-3).

3.3.5 Cells expressing fusion show increased anchorage-independent growth

To determine if the fusion induced malignant transformation, the 6 fusion cell lines and control cells (original and transduced with Cas9) were tested in a soft agar colony formation assay, and the extent of colony formation was assessed after 2 weeks of growth in agarose. We observed a significant increase in colony formation potential in several of the cells expressing the PRKAR1A-RET rearrangement: HATB_3, HBTB_1, HBTB_2 and HBTB_3 as compared to control cells (Fig. 3.4).

3.3.6 Transformed cells with fusion exhibit differential regulation

We next determined protein expression and signaling changes in cells that displayed anchorage-independent growth, which has been reported to correlate with tumorigenesis. To do so, we picked colonies and subsequently expanded them in culture for 2 passages to generate sufficient protein for proteomic and phosphoproteomic analyses. In order to separate fusion-specific effects from growth in soft agar, control cells (Nthy-ori 3-1) having formed smaller but clear colonies underwent parallel sample handling (Fig. 3.2).

As was the case for the non-transformed cells, we sought to compare regulated proteins in our data with proteins previously reported as relevant to thyroid cancers. We identified regulation of TPR in fusion-expressing colonies. We also identified modulation of HMGA1 expression in transformed cells. Previous studies have shown that elevated HGMA1 and HGMA2 levels are required for transformation of thyroid cells. [18]. Also known to be involved specifically in thyroid cancer and showing differential phosphosite regulation in transformed cells were KTN1, ERC1 and PCM1. As is the case for KTN1 and ERC1, PCM1 has been also been found rearranged in thyroid cancers, although in this case to RET [19].

3.3.7 Cells homozygous for fusion increasingly enrich for cancer hallmarks

During single cell cloning we noted that two of the six clones (HBTB_1 and HBTB_3) had a shorter doubling time in culture. These two clones also showed the largest increase in their ability for anchorage-independent growth compared to control cells. Although sequencing of the clones confirmed the

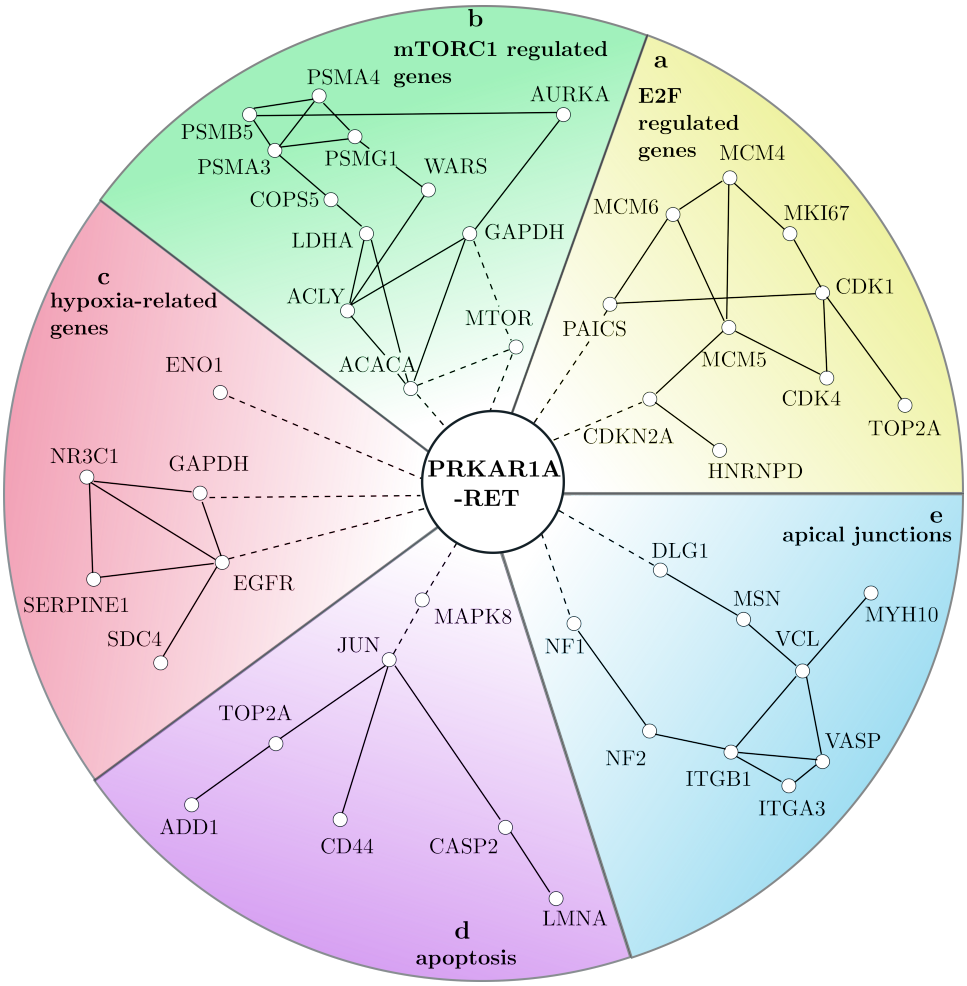


FIGURE 3.3: Strongly enriched cancer hallmarks and possible mechanisms in which the PRKAR1A-RET fusion could exert its effects. Solid lines link network nodes that have been observed as regulated by mass spectrometry. Dashed lines correspond to edges involving not observed proteins. **a** and **b** show the genes regulated in non transformed cells upon fusion expression associated with upregulation of E2F and mTORC1, respectively. **c** depicts the genes upregulated under hypoxic conditions that were regulated in non transformed fusion-harboring cells. **d** Phosphoproteins regulated in transformed cells were enriched in genes mediating apoptosis. **e** Genes encoding for the apical junction complex were found to be enriched specifically in transformed cells.

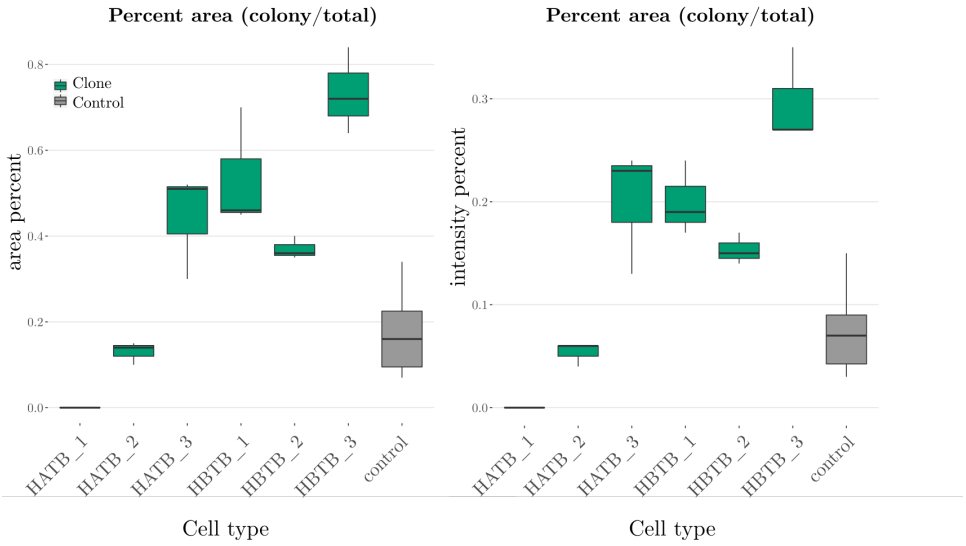


FIGURE 3.4: Soft agar colony formation assays to assess anchorage-independent growth of fusion-expressing cells. Left to right: Percent intensity and percent area of the assay area covered by colonies, respectively. Box plots summarize data obtained from triplicate repeats where control cells did not express the PRKAR1A-RET fusion.

presence of the fusion in all selected cases, we noted an increase in the noise level for the four clones which did not show a decreased doubling time (Supp. Fig. 3.3). This could be due to the selection of a mixed population at the moment of single-cell cloning. While the HBTB_1 and HBTB_3 clones are clearly homozygous for the translocation, we can only speculate that the other clones might be cases of heterozygosity.

We therefore wished to assess the processes affected in the cells with potentially the most aggressive fusion phenotype and retained HBTB_1 and HBTB_3 as representative fusion-containing cells and colonies of these clones. The 743 regulated proteins in this subset enriched for similar processes as the non-transformed cells. As was the case prior to transformation, the most enriched hallmarks were MYC targets (FDR = 1.12x10-41), E2F targets (FDR =1.05x10-25), oxidative phosphorylation (FDR = 1.82x10-23), upregulated proteins related to mTOR signaling (FDR = 3.16x10-20), G2/M checkpoint (FDR = 2.69x10-13) and protein secretion (FDR = 5.21x10-10).

In contrast to the non-transformed cells, we identified an enrichment for genes upregulated by reactive oxygen species (ROS) (FDR = 1.56x10-11). *In vivo*, inflammatory cells can release ROS, which are deleterious for neighboring cancer cells. Genes regulated as a consequence of the presence of reactive oxygen species are known to drive cells toward a malignant phenotype [20]. We also noted proteins associated with the formation of apical junctions (FDR = 1.23x10-10) to be enriched. This is likely a key shift allowing cells to exhibit anchorage-independent growth and thus for their transformation. Indeed, epithelial cancers often display loss of cell-cell adhesion and cell polarity in order to gain in plasticity, leading to increased invasiveness [21].

Interestingly we found notably more altered phosphorylation in the colonies than we did the fusion expressing cells. We identified 404 significantly modulated phospho-proteins compared to the ThyCas9 cells. These were enriched

for MYC targets (FDR = 4.36x10-39), E2F targets (FDR = 1.59x10-26), G2/M checkpoint (FDR=5.81x10-19), and upregulated proteins related to mTOR signaling (FDR =3.74x10-10). As was the case for differentially expressed proteins, modulated phosphosites enriched for proteins associated with the formation of apical junctions (FDR=2.51x10-7), suggesting that this hallmark is specific to the transformed state. We noted that regulated phosphoproteins in the colonies carrying the fusion were also enriched in apoptosis, which is often suppressed in cancers (FDR=2.28x10-9).

3.4 Discussion

To the best of our knowledge, this study represents the most extensive sampling of the proteomic and phoshoproteomic network changes upon endogenous expression of a chimeric gene found in solid tumors. We identified and quantified many changes occurring in cells following the endogenous expression of the fusion. We confirmed the regulation of several genes known to be involved in thyroid cancers. Furthermore, our analysis showed that cells harboring the induced fusion enriched for processes essential for tumorigenesis. Notably, we identified enrichment for genes involved in regulating the cell cycle, critical for uncontrolled proliferation and the most prominent cancer hallmark. We also noted increased regulation of MYC and mTOR targets, whose implication in a large number of cancers is well-established.

After demonstrating that the PRKAR1A-RET fusion caused increased malignant transformation in immortalized thyroid cells, we showed that the regulated proteins and phosphosites further enriched for cancer hallmarks. We observed a large overlap of enriched properties compared to non transformed cells, albeit with higher enrichment in transformed cells. Strikingly, we observed the appearance of apical junctions as a marker for cells in a transformed state following fusion expression. We speculate that this is an important mechanism by which cells may acquire changes related to adhesion and migration, and potentially invade their surroundings during tumorigenesis.

Rearrangements have now been identified in myriad different solid tumor cancers and have been shown to be able to confer cells harboring them increased malignant behavior. However with the exception of studies investigating blood-related diseases, there are few experimental systems-wide studies of the effect of chromosomal translocations. *In silico* predictions of rewired protein-protein interactions upon fusion expression suggest that there are commonalities in the general signaling mechanisms of fusions that lead to cancer [14]. The overlap of our regulated proteins with the predictions of affected interactors was therefore encouraging, and the work presented here is a direct validation of several of these predictions.

We believe it is of the utmost importance to study the effects of such perturbations in a context as close as possible to the physiological setting. Indeed, many studies investigating RET rearrangements in thyroid cancer were performed in non-thyroid cells. Moreover, the majority use a viral or plasmid vector to introduce the fusion construct in cells. These are generally expressed under strong promoters and thus might lead to an expression of the translocation which does not reflect *in vivo* conditions. While these have provided critical insights into the targets and mechanisms of the fusion, the

results are likely confounded by the aforementioned limitations. Here we have therefore created a model to study PRKAR1A-RET expressed at endogenous levels in a non-transformed thyroid cell line, mimicking conditions in which the fusion would arise.

To consolidate the findings presented here, we propose to induce the fusion in primary cells and assess its potential for full transformation. Beyond gleaning an improved understanding of the mechanisms of the expressed fusion product, the ultimate aim would be to determine the role of the kinase fusion during transformation to assess its use as a druggable target for the treatment of cancer.

3.5 Author contributions

F.V., C.D.S. and C.H. designed and produced the Cas9 expressing thyroid cells and fusion-expressing cells. F.V. conducted the mass spectrometry experiments. F.V. and G.S. designed and performed the colony formation assays. F.V. and X.R. analyzed the data. F.V., C.D.S., J.L. and R.L. conceived the project. J.L. gave conceptual advice. R.L. supervised the project. F.V., J.L. and R.L. wrote the manuscript. All authors edited the manuscript.

3.6 Acknowledgements

This work was funded by a Lundbeck Foundation Fellowship awarded to R.L., an European Research Council starter grant awarded to R.L. under the European Union?s Seventh Framework Programme and the Innovation Fund Denmark Grand Solutions project MorphoMap (1311-00010B) Funding for F.V. was provided by the Technical University of Denmark.

References

[1] N. Stransky, E. Cerami, S. Schalm, J. L. Kim, and C. Lengauer, "The landscape of kinase fusions in cancer.", *Nature Communications*, no. 4846, 2014.

[2] S. Medves, L. A. Noël, C. P. Montano-Almendras, R. I. Albu, H. Schoemans, S. N. Constantinescu, and J. B. Demoulin, "Multiple oligomerization domains of KANK1-PDGFRβ are required for JAK2-independent hematopoietic cell proliferation and signaling via STAT5 and ERK.", *Haematologica*, no. 10, pp. 1406–14, 2011.

[3] S. Reckel, R. Hamelin, S. Georgeon, F. Armand, Q. Jolliet, D. Chiappe, M. Moniatte, and O. Hantschel, "Differential signaling networks of BcrAbl p210 and p190 kinases in leukemia cells defined by functional proteomics.", *Leukemia*, no. 7, pp. 1502–1512, 2017.

[4] J. Cutler, R. Tahir, S. Sreenivasamurthy, C. Mitchell, S. Renuse, R. Nirujogi, A. Patil, M. Heydarian, X. Wong, X. Wu, T.-C. Huang, M.-S. Kim, K. Reddy, and A. Pandey, "Differential signaling through p190 and p210 BCR-ABL fusion proteins revealed by interactome and phosphoproteome analysis.", *Leukemia*, no. 7, pp. 1513–1524, 2017.

[5] I. Bongarzone, M. Butti, S. Coronelli, M. Borrello, M. Santoro, P. Mondellini, S. Pilotti, A. Fusco, G. Della Porta, and M. Pierotti, "Frequent activation of ret protooncogene by fusion with a new activating gene in papillary thyroid carcinomas.", *Cancer Research*, vol. 54, no. 11, pp. 244–50, 1994.

[6] V. Roukos, T. C. Voss, C. K. Schmidt, S. Lee, D. Wangsa, and T. Misteli, "Spatial Dynamics of Chromosome Translocations in Living Cells.", *Science*, vol. 341, no. 6146, pp. 660–4, 2013.

[7] E. Mercalli, S. Ghizzoni, E. Arighi, L. Alberti, R. Sangregorio, M. T. Radice, M. L. Gishizky, M. A. Pierotti, and M. G. Borrello, "Key role of Shc signaling in the transforming pathway triggered by Ret-ptc2 oncoprotein.", *Oncogene*, vol. 20, no. 27, pp. 3475–85, 2001.

[8] D. Maddalo, E. Manchado, C. P. Concepcion, C. Bonetti, J. A. Vidigal, Y.-C. Han, P. Ogrodowski, A. Crippa, N. Rekhtman, E. De Stanchina, S. W. Lowe, and A. Ventura, "In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system.", *Nature*, vol. 516, no. 7531, pp. 423–7, 2014.

[9] R. Torres, M. C. Martin, A. Garcia, J. C. Cigudosa, J. C. Ramirez, and S. Rodriguez-Perales, "Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPRCas9 system.", *Nature Communications*, vol. 5, no. 3964, 2014.

[10] P. S. Choi and M. Meyerson, "Targeted genomic rearrangements using CRISPR/Cas technology.", *Nature Communications*, vol. 5, 2014.

[11] A. Shin, V. Freedman, R. Risser, and R. Pollack, "Tumorigenicity of virus-transformed cells in nude mice is correlated specifically with anchorage independent growth in vitro.", *Proc Natl Acad Sci U S A.*, no. 11, pp. 4435–4439, 1975.

[12] N. R. Lemoine, E. S. Mayall, T. Jones, D. Sheer, S. Mcdermid, P. Kendall-Taylor, and D. Wynford-Thomas, "Characterisation of human thyroid epithelial cells immortalised in vitro by simian virus 40 DNA transfection.", *Br. J. Cancer*, vol. 60, pp. 897–903, 1989.

[13] P. A. Futreal, "A Census of Human Cancer Genes.", *Nature Reviews Cancer*, vol. 4, no. 3, pp. 177–83, 2004.

[14] M. Frenkel-Morgenstern, A. Gorohovski, S. Tagore, V. Sekar, M. Vazquez, and A. Valencia, "ChiPPI: a novel method for mapping chimeric proteinprotein interactions uncovers selection principles of protein fusion events in cancer.", *Nucleic Acids Research*, vol. 45, no. 12, pp. 7094–7105, 2017.

[15] A. Liberzon, C. Birger, H. Thorvaldsdottir, M. Ghandi, and P. Mesirov, "The Molecular Signatures Database Hallmark Gene Set Collection.", *Cell Systems*, no. 6, pp. 417–425, 2015.

[16] E. Souza, A. Ferreira, and D. Carvalho, "The mTOR protein as a target in thyroid cancer.", *Expert opinion on therapeutic targets*, no. 9, pp. 1099–112, 2011.

[17] D. Buehler, H. Hardin, W. Shan, C. Montemayor-Garcia, P. Rush, S. Asioli, H. Chem, and R. Lloyd, "Expression of epithelial-mesenchymal transition regulators SNAI2 and TWIST1 in thyroid carcinomas.", *Mod Pathol.*, no. 2, pp. 54–61, 2012.

[18] M. Berlingieri, G. Pierantoni, V. Giancotti, M. Santoro, and A. Fusco, "Thyroid cell transformation requires the expression of the HMGA1 proteins.", *Oncogene*, no. 19, pp. 2971–80, 2002.

[19] R. Corvi, N. Berger, R. Balczon, and G. Romeo, "RET/PCM-1: a novel fusion gene in papillary thyroid carcinoma.", *Oncogene*, no. 37, pp. 4236–42, 2000.

[20] D. Hanahan and R. A. Weinberg, "Hallmarks of Cancer: The Next Generation.", *Cell*, no. 5, pp. 646–74, 2011.

[21] D. Coradini, C. Casarsa, and S. Oriana, "Epithelial cell polarity and tumorigenesis: new perspectives for cancer detection and treatment.", *Acta pharmacologica Sinica*, no. 5, pp. 552–64, 2011.

[22] S. A. Forbes, G. Bhamra, S. Bamford, E. Dawson, C. Kok, J. Clements, A. Menzies, J. W. Teague, P. A. Futreal, and M. R. Stratton, "The catalogue of somatic mutations in cancer (COSMIC)", *Current Protocols in Human Genetics*, 2008.

[23] M. Frenkel-Morgenstern, A. Gorohovski, V. Lacroix, M. Rogers, K. Ibanez, C. Boullosa, E. A. Leon, A. Ben-Hur, and A. Valencia, "ChiTaRS: A database of human, mouse and fruit fly chimeric transcripts and RNA-sequencing data.", *Nucleic Acids Research*, pp. D142–51, 2013.

[24] S. Q. Tsai, Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, and A. J. Iafrate, "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.", *Nature Biotechnology*, no. 2, pp. 1–3, 2014.

[25] S. Borowicz, M. Van Scoyk, S. Avasarala, M. K. Karuppusamy Rathinam, J. Tauler, R. K. Bikkavilli, and R. A. Winn, "The soft agar colony formation assay.", *Journal of visualized experiments : JoVE*, no. 92, e51998, Oct. 2014.

[26] C. Guzmá, M. Bagga, A. Kaur, J. Westermarck, and D. Abankwa, "ColonyArea: An ImageJ Plugin to Automatically Quantify Colony Formation in Clonogenic Assays.", *PLoS ONE*, vol. 9, no. 3, e92444, 2014.

[27] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: A flexible trimmer for Illumina sequence data.", *Bioinformatics*, no. 15, pp. 2114–20, 2014.

[28] J. Li, U. Rix, B. Fang, Y. Bai, A. Edwards, J. Colinge, K. L. Bennett, J. Gao, L. Song, S. Eschrich, G. Superti-Furga, J. Koomen, and E. B. Haura, "A chemical and phosphoproteomic characterization of dasatinib action in lung cancer.", *Nature Chemical Biology*, vol. 6, no. 4, pp. 291–9, 2010.

[29] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo, "The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.", *Genome Research*, no. 9, pp. 1297–303, 2010.

[30] M. A. Depristo, E. Banks, R. E. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, A. Mckenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly, "A framework for variation discovery and genotyping using next-generation DNA sequencing data.", *Nature Genetics*, no. 5, pp. 491–8, 2011.

[31] W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, and F. Cunningham, "The Ensembl Variant Effect Predictor.", *Genome Biology*, no. 1, p. 122, 2016.

[32] S. Tyanova, T. Temu, and J. Cox, "The MaxQuant computational platform for mass spectrometrybased shotgun proteomics.", *Nature Protocols*, vol. 11, no. 12, pp. 2301–2319, 2016.

[33] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016.

[34] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik, *Cluster: Cluster analysis basics and extensions*, 2016.

[35] S. Durinck, P. Spellman, E. Birney, and W. Huber, "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt.", *Nature Protocols*, pp. 1184–1191, 2009.

[36] Wickham and Hadley, "Reshaping data with the reshape package", *Journal of Statistical Software*, vol. 21, no. 12, 2007.

[37] ——, "Ggplot2: Elegant graphics for data analysis", 2009.

[38] H. Wickham and R. Francois, *Dplyr: A grammar of data manipulation*, R package version 0.5.0, 2016.

3.7 Materials and Methods

3.7.1 Cell lines and vectors

SV40 immortalized primary cells (n-thy-ori) from thyroid follicular epithelial tissue were purchased from Sigma and cultured in RPMI 1640 medium supplemented with 2 mM glutamine, 10% FBS, 1% antibiotic-antimycotic solution (all Gibco) in a humidified environment with 5% CO2. HEK-293T cells were cultured in DMEM, supplemented with 10% PBS and 1% penicillin/streptomycin. pCMV-VSV-G, psPAX2, pCW-Cas9 Dox-inducible lentiviral vector and pLX-sgRNA lentiviral vector containing AAVS1-targeting sgRNA from Addgene.

3.7.2 Viability assay

Cells were seeded at 4e5 cells/well in growth medium in 6-well tissue culture plates. Media was substituted with increasing concentrations of puromycin the following day and was maintained for 3 days, after which growth media was given for 3 days to allow potential recovery/resistance. Cells were washed 3 times with PBS and stained with a Crystal Violet solution for 30 minutes followed by several PBS washes.

3.7.3 Vector construction

The pLX-sgRNA vector was digested with Xba1, Nde1 for 2 hrs at 37C, ran on a 0.5% agarose gel followed by gel extraction and purification. The DNA digest was further used to clone a new vector allowing for facile insertion of sgRNA oligonucleotides via Gibson reactions. In order to select for successful transduction of the second sgRNA, the vector was further modified to replace the blasticidin resistance with hygromycin.

3.7.4 Determination of targeted breakpoint location

The empirically observed breakpoint of the translocation event was verified using COSMIC [22] and ChiTars [23] and the introns selected by following the principle of last observed- and first observed exons given several documented cases of the same translocation, with varying numbers of exons detected in the partner genes. Only one combination of introns targeted for CRISPR-Cas9-mediated cleavage was retained, using the intron following the last seen exon of the head partner and the intron preceding the first observed exon of the tail moiety.

3.7.5 Lentiviral sgRNA design

The tool developed by Doench *et al.* was used to score all SpCas9 Protospacer Adjacent Motif (PAM) sites within selected introns. The two top-scoring guide sequences were retained for each translocation partner. A 'G' was prepended to the guide sequence (GX20 sgRNA), with the benefit of diminishing off-targets at the cost of cutting efficiency [110]. Top-scoring guide sequences were verified by submission to Zhang *et al.* web-tool. Selected sgRNAs were verified with updated off-target prediction algorithm GUIDE-seq [24].

3.7.6 HEK-293 infection

2e6 HEK-293T packaging cells were seeded in a 75cm2 cell culture flask in growth medium without antibiotics. The following day, cells were transfected with pCMV-VSV-G, psPAX2, and pCW-Cas9 vectors using MirusBio TransIT-LT1 transfection reagent, and after 24 hours media containing high-BSA was substituted. 24 hours later media was harvested and residual cells were removed by filtration.

3.7.7 Viral pCW-Cas9 transduction and titration

Nthy-ori 3-1 cells were seeded at 300'000 cells/well in a 6-well plate. The following day, protamine sulfate was added to flasks to a final 8 μg/ml and cells were transduced with the Cas9 construct with increasing volumes of virus-containing harvested media : 0, 2, 5, 10, 20, 50, 75, 100, 150, 200, 250 μL. After overnight incubation, growth media was replaced.

Cells with increasing viral titers were seeded in 96-well tissue culture plates at 5000 and 10000 cells/well in 200 μL of media. Growth media was replaced with medium supplemented with 0.8 μg/mL puromycin the following morning. Alamar Blue was added to the wells and % viable cells was measured at 570 nm after 5 hours and after 24 hours. The lowest viral

titer at which cells were viable post selection was chosen to move forward, and cells were further maintained in puromycin-containing media.

3.7.8 Dosing of doxycylin as Cas9-inducing agent

4e5 Thy-Cas9 cells were sparsely seeded in 10cm dishes and the following morning doxycycline was added at 1 μg/mL in each dish (previous assay showed 1μg/mL to show sufficient activity). Cells were harvested every 24 hours in duplicates or triplicates in Laemmli sample buffer followed by Western Blot analysis. Cultures without doxycycline were maintained and lysed in parallel as negative controls. Analysis by immunoblot showed peak Cas9 expression at 72 hours followed by a decline to barely detectable activity after one week.

3.7.9 Viral modified pLX-sgRNA transduction

Media containing viral sgRNA vectors was harvested as previously described for each individual sgRNA. 1e6 Thy-Cas9 cells were seeded per 25 cm^2 cell culture plate. The following morning harvested media was added to growth media in increasing volumes, at equal volumes between the two sgRNAs of a given combination. Cells were split twice and selected for 14 days in media supplemented with 0.8 μg/mL puromycin, 10 μg/mL blasticidin or 200 μg/mL hygromycin. The condition with lowest viral volume mediating resistance was retained for further experiments.

3.7.10 Cas9 expression induction

Doxycycline at a final concentration of 1 μg/mL was administered for 1 week to 4e5 cells seeded in 100mm tissue culture dish. Doxycycline-containing media was refreshed after 4 days. Immediately thereafter individual clones were generated via single cell cloning using clonal rings and were expanded (96-well was named "p1 after single cell cloning").

3.7.11 Translocation detection PCR

RNA was extracted using Qiagen RNA extraction kit from cells at various passages following single cell cloning. Clones were screened for translocations and the cDNA breakpoint determined by RT-PCR in a 96-well AB Veriti Fast Thermal Cycler using the Quantitect SYBR green RT-PCR kit (Qiagen) at the following the manufacturer's guidelines and using an annealing temperature of 56C for 30 s, and extension at 72C for 60 s, and a final extension of 5 min at 72C. The forward primer was AGACAATGGCCGCTTTAGCCA and the reverse primer was CAGGGAGCCGTATTTGGCGT. Sanger sequencing was used to confirm sequence of bands at correct weight 3.3.

3.7.12 Anchorage-independent growth assay

Transformation was verified using the soft agar colony formation assay [25]. 6-well tissue culture plates were coated with growth medium containing 1% agarose, and triplicate dishes were inoculated with 10'000 or 2'000 cells suspended in RPMI-1640 containing 0.8% agarose and 15% FBS, for counting and colony picking, respectively. Colony formation was assessed after 2

weeks. Image analysis was done using ImageJ, with plugins ColonyArea [26] and ColonyCounter (https://imagej.nih.gov/ij/plugins/colony-counter.html).

3.7.13 Mass spectrometry lysis and sample preparation

Protein was harvested by lysing a 150 mm dish at 80-90% confluency in 1 mL modified RIPA buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 1 mM EDTA) supplemented with phosphatase inhibitors (5 mM þ-glycerophosphate, 5 mM NaF, 1 mM Na_3VO_4) and protease inhibitors (Roche cOmplete ULTRA Tablets, EDTA-free, #05892791001), sonicated, and spun down for 20 min at 4300G. Ice-cold acetone was added to the supernatant to achieve a final concentration of 80% acetone, and protein was left to precipitate overnight at -20°. Precipitated protein was pelleted by centrifugation at 1800G for 5 min and solubilized in 6 M urea, 2 M thiourea, 10 mM HEPES pH 8.0. Protein was quantified using the Bradford assay and 650 μg of each sample were reduced with 1 mM DTT, alkylated with 5 mM ClAA, and digested with endopeptidase Lys-C (1:200 v/v, Wako #129-02541) for 3 hours. Samples were diluted to 1 mg/mL protein using 50 mM ammonium bicarbonate and incubated overnight with trypsin (1:200 v/v, Sigma #T6567). Digested samples were acidified and urea removed using Sep-Pak tC18 96-well plate (Waters #186002320) C18 cartridges. An aliquot of 50μg of eluted peptides was set aside for proteome analysis as well as peptide quantitation (Pierce quantitative colorimetric peptide assay, Thermo #23275) to equalize injection load for label-free LC-MS/MS analysis. The remainder was lyophilized, aliquots of 200μg made up in 1 M glycolic acid, 80% ACN, 5% TFA, and enriched for phosphopeptides using MagReSyn Ti-IMAC beads (#MR-TIM010 ReSyn Biosciences) following the manufacturer's protocol. Enriched peptides were subjected to a final C18 clean-up prior to data acquisition.

3.7.14 MS data acquisition

All spectra were acquired on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) operated in data-dependent mode coupled to an EASY-nLC 1200 (Thermo Fisher Scientific) liquid chromatography pump and separated on a 50cm reversed phase column (Thermo, PepMap RSLC C18, 2μM, 100Å, 75μmx50cm). Proteome samples (non-enriched) were eluted over a linear gradient ranging from 0-11% ACN over 70 min, 11-20% ACN for 80 min, 21-30% ACN for 50 min, 31-48% ACN for 30 min, followed by 76% ACN for the final 10 min with a flow rate of 250 nL/min. Phosphopeptide- or phosphotyrosine enriched samples were eluted over a linear gradient ranging from 0-18% ACN over 195 min, 18-26% ACN for 30min, 26-76% ACN for 10min, followed by 76% ACN for the final 5 min with a flow rate of 250 nL/min.

Survey full scan MS spectra were acquired in the Orbitrap at a resolution of 120000 from m/z 350-2000, AGC target of 4e5 ions, and maximum injection time of 20 ms. Precursors were filtered based on charge state (≥ 2) and monoisotopic peak assignment, and dynamic exclusion was applied for 45s. A decision tree method allowed fragmentation for ITMS2 via ETD or HCD, depending on charge state and m/z. Precursor ions were isolated with the

quadrupole set to an isolation width of 1.6 m/z. MS2 spectra fragmented by ETD and HCD (35% collision energy) were acquired in the ion trap with an AGC target of 1e4. Maximum injection time for HCD and ETD was 200 ms for phosphopeptide-enriched samples, and 80ms for proteome samples.

3.7.15 Fasta file generation

DNA was extracted from the MDA-MB 231 cell line at 80% confluence using the Qiagen QIAamp DNA Mini kit, according to the manufacturers' instructions. Sequencing was performed by Beckman Coulter Genomics with a target average exome coverage of 89x. We filtered the reads with Trimmomatic [27] with the following parameters: headcrop = 3, minlen = 30, trailing = 3. We aligned the trimmed reads to the hg19 reference genome with BWA [28], applied GATK [29] base quality score recalibration, indel realignment, duplicate removal, and performed SNP and INDEL discovery and genotyping across all samples simultaneously using standard hard filtering parameters according to GATK Best Practices recommendations [30]. We used the Ensembl Variant Effect Predictor (VEP) [31] with Ensembl v. 75 to predict the effect of the mutations on the protein sequence. In order to find non-reference, mutated peptides in the MS data, we increased the search Fasta file with mutations affecting the protein sequence detected by WES with a high sensitivity filter: QD $\leq$ 1.5, FS $\geq$ 60, MQ $\geq$ 40, MQRankSum $\leq$ -12.5, ReadPosRankSum $\leq$ -8.0, and DP at least 5 / sample on average.

3.7.16 Mass spectrometry identification and quantification

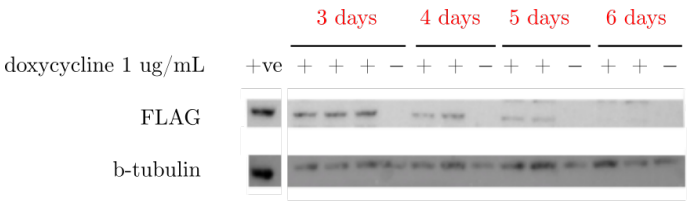
All label-free raw data were searched using MaxQuant v1.5.1.2 against the human Ensembl database (v88), the mutant proteins from exome sequencing as well as the sequence of the fusion product [32]. Default settings were used with the exception of a minimum peptide length of 6 amino acids. STY phosphorylation was added as a variable modification. An FDR of 0.01 was applied at the level of proteins, peptides and modifications. Phosphosites were filtered for a localization score of at least 0.9.

3.7.17 Data analysis

All data analysis was carried out in R, with additional packages [33]–[38]. Label-free quantification (LFQ) was performed based on extracted ion chromatograms (XIC) for identified peptides after retention time alignment by MaxQuant. XICs were normalized per peptide. Log ratios used throughout were taken to the ThyCas9 control.

3.7.18 Markov Chain Monte Carlo

A likelihood function for each concentration c or phosphorylation occupancy o ratio was calculated as follows:



SUPPLEMENTARY FIGURE 3.1: FLAG-tagged Cas9 expression optimization. The lowest dose of doxycycline showing sufficient Cas9 expression was selected for further experiments. Expression of Cas9 peaks 72 hours after induction and decreases thereafter, indicating requirement to replenish doxycyclin-supplemented media for induction periods longer than 3 days.

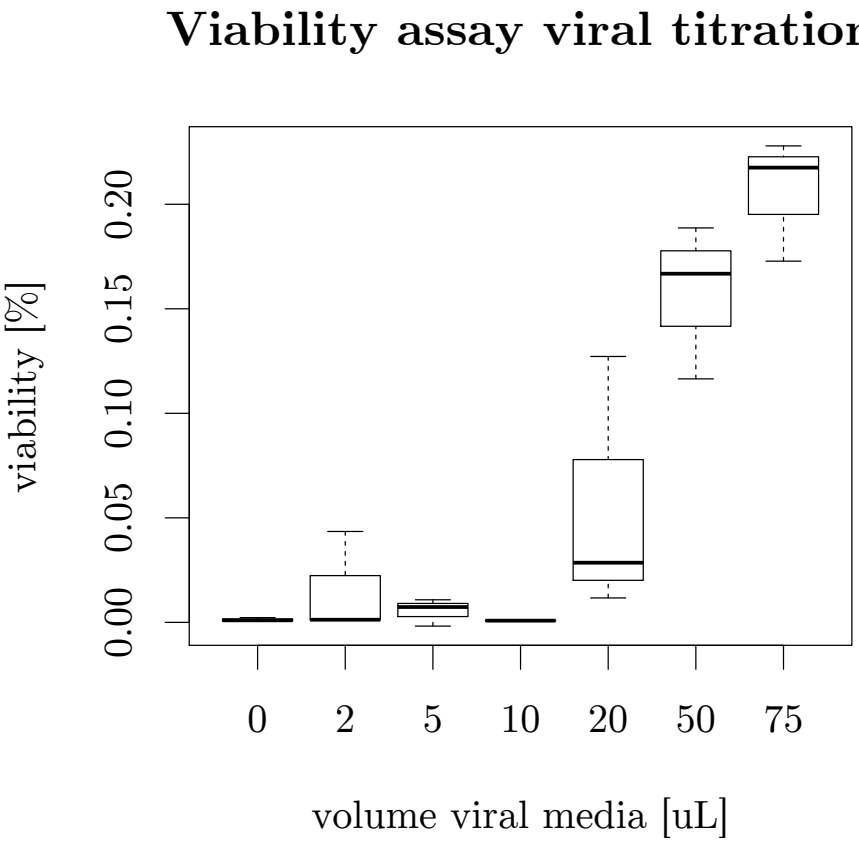
$$
\begin{aligned}
L(c, \bar{o}, \bar{o}') &= \log P(\bar{z} | \bar{\mu}(c, \bar{o}, \bar{o}')) \\
&= \sum_{i=1}^{I} \Bigg(\tilde{\nu}_i \log \Bigg[1 + \gamma_i \Big(\bar{x}_i - c - \\
&\qquad \sum_s I_{is} \left[\log \left(\frac{1 - o'_s}{1 - o_s} \right) + t_{is} \left(\log \left(\frac{o'_s}{1 - o'_s} \right) - \log \left(\frac{o_s}{1 - o_s} \right) \right) \right] \Big)^2 \Bigg] + \\
&\qquad \frac{1}{2} \log(\gamma_i) \Bigg)
\end{aligned}
\tag{3.1}
$$

where $\hat{x}$ is a vector of the peptide ratios mapped to the protein, I and t are indicator matrices of size $i \times s$ reporting whether peptide i spans phosphosite s (I) or is modified on site s (t), and ν and γ are Bayesian variance parameters established by fitting Beta distributions by binned absolute log ratios. We sampled this likelihood for each protein separately with Bayesian Markov Chain Monte Carlo (MCMC) with Jeffrey priors on o and o' with $\alpha_1 = \alpha_2 = \frac{1}{2}$, and an exponential prior with $\lambda = 2$ on c.

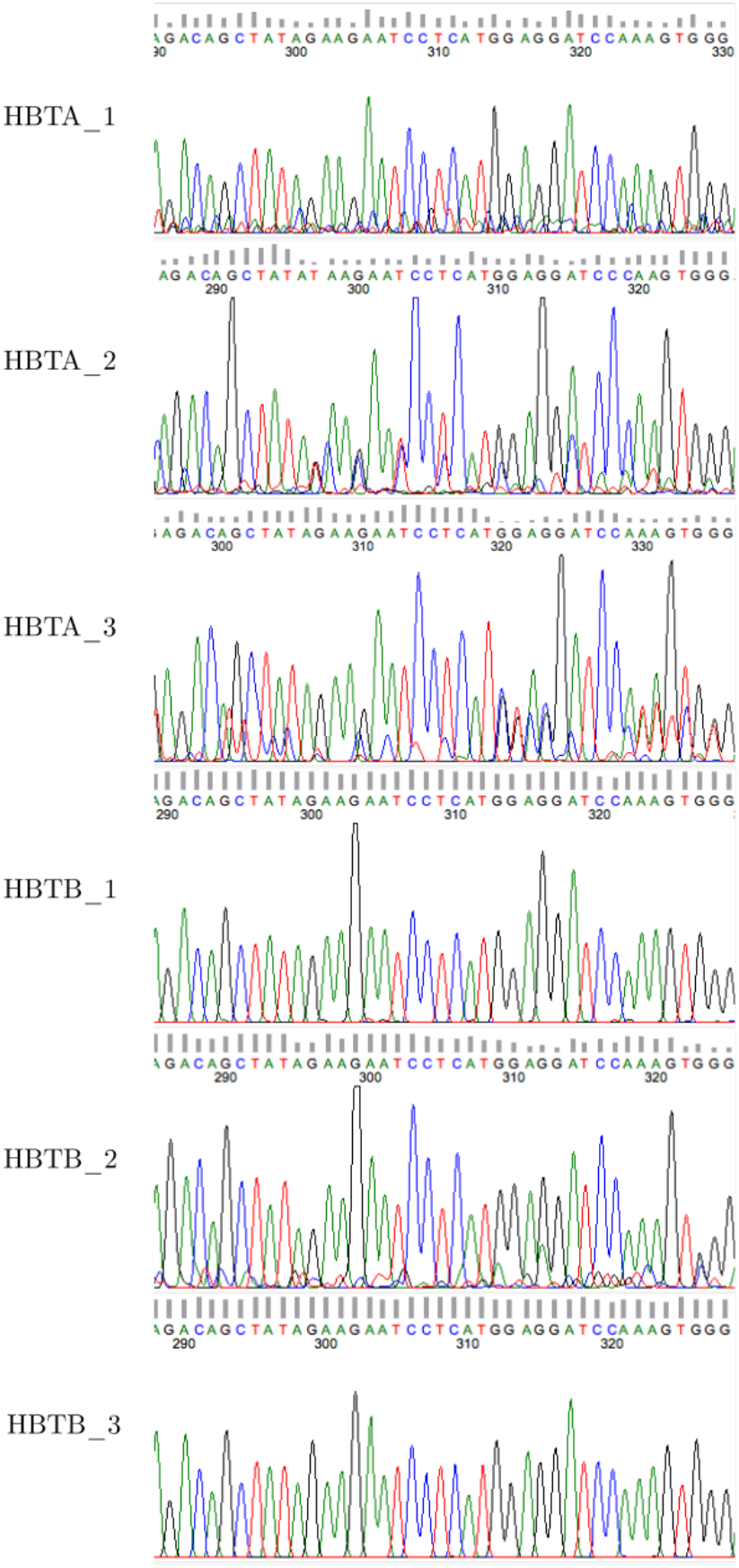
Finally we calculated the means and standard deviations of the chains, resulting in means and standard errors for each protein and phospho-site occupancy. We calculated phosphosite ratios for a given sample to the control as $log(o.sample/(1 - o.sample)) - log(o.reference/(1 - o.reference))$.

3.7.19 Functional enrichment analyses

The hallmark gene sets of the Molecular Signatures Database (MSigDB) were used to determine enriched hallmarks.



Supplementary Figure 3.2: Viability assay for the determination of viral titer for lentiviral transduction in n-thy-ori cells.



Supplementary Figure 3.3: cDNA PCR to identify expression of fusion product in single cell-derived clones. The red rectangle denotes the overlapping region between the gene pair, PRKAR1A and RET.

Chapter 4

Proteogenomic analysis of dynamic cell signaling responses to wounding

Franziska Voellmy[1,2,7], Gaye Saginc[1,7], Mathias Engel[1,3,7], Xavier Robin[1], Tom Altenburg[4], Xiaokang Lun[5], Bernd Bodenmiller[5], James Longden[1]*, Rune Linding[1,6]*

[1] Biotech Research and Innovation Centre, University of Copenhagen, Copenhagen, 2200, Denmark

[2] Center for Biosustainability, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark

[3] Niels Bohr Institute, University of Copenhagen, Copenhagen, 2200, Denmark

[4] Institute of Biology, Humboldt-Universität zu Berlin, Berlin, 10115, Germany.

[5] Institute of Molecular Life Sciences, University of Zurich, Zurich, CH-8057, Switzerland

[6] Lead contact

[7] These authors contributed equally

* correspondence to linding@lindinglab.org or longden@lindinglab.org

4.1 Summary

The acquisition of an invasive phenotype is the first step in the development of metastatic potential, and we therefore aimed to identify signaling processes implicated in migrating cells. We used high-throughput LC-MS/MS to quantify the modulation of phosphorylation in triple negative breast cancer cells upon wounding. The tyrosine and serine/threonine phosphoproteome was sampled at frequent and extremely early time points(ranging from several seconds up to one hour) in order to generate detailed temporal signaling

profiles per phosphorylated residue. A large portion of the quantified phosphoproteome was found to be regulated at the earliest measured time points (sub-minute range), showed an increase in tyrosine phosphorylation, and specific sites were found to be enriched for cell motility. The increased ability of proteins harboring modulated phosphosites to affect migration was confirmed using siRNA-mediated knockdown, showing a change in cell migration velocity. This study identifies a number of currently unannotated phosphorylation sites potentially involved in wound healing and migration, suggesting their potential as novel class of metastatic-reducing drug targets.

4.2 Keywords

Proteomics, wound healing, cell signaling, phosphorylation, cancer

4.3 Significance

Current specific cancer therapies are developed to act upon kinases by competing for their ATP-binding site, thus decreasing their ability to phosphorylate their substrates and thus interrupt signal transduction. We identify distinct phosphorylation sites within a single protein to be implicated in different cellular processes. This suggests that, although sometimes successful, targeting the active site of kinases is likely to lack specificity in terms of the targeted process. In order to target cell migration and therefore possibly reduce metastatic burden in cancer patients, we believe that rendering phosphosites inactive could be sufficient to control disease. Over time, as our knowledge of the precise cellular events in which phosphorylation sites are involved improves, this information may be used in research to develop more intricate models of cellular signaling, as well as for the development of novel therapeutics.

4.4 Highlights

- Breast cancer cell monolayer wounding resulted in immediate signaling response
- Very early temporal signaling dynamics showed an increase in tyrosine phosphorylation
- Modulated phosphorylation signaling was associated with migration
- Novel targets for anti-metastatic therapies identifed by modulated signaling peptides

4.5 Introduction

"Tumors are wounds that do not heal" [1]. A link between cancer and wound healing processes has been thought to exist since observations of similarities between tumor stroma and tissue at the site of healing skin wounds, suggesting that common molecular mechanisms might underlie the two events. Cellular processes responsible for wound healing are complex and tightly modulated in order to restore tissue at the site of injury. The wound-healing

process and carcinogenesis share the following processes: inflammation, cell proliferation, migration and invasion. Although certain cancers may activate this latent wound-healing program to modify connective tissue as needed for their sustained growth, they lack the regulation of the normal wound repair process and as a consequence uninhibited cell proliferation and invasion ensue.

Inflammation was found to be required for wound-related tumor formation at sites of injury [2]. Indeed, several cancers are known to arise directly as a result of a chronic inflammatory response, such as that caused by Helicobacter pylori infections and irritable bowel syndrome. [3]. Inflammation is a hallmark of cancer and is critical for disease progression [4], [5], providing a framework for sustenance of the inflammatory response as well as generating reactive oxygen species (ROS) which in a cancer setting may affect DNA repair mechanisms and are thought to contribute to enhanced malignant transformation [6]. Following wounding, cells at the edge of the wound exhibit directed migration in order to fill the void at the site of injury, providing another parallel to cancerous cells.

One of the earliest events in the metastatic process is the invasion of cells located in the primary tumor into their environment. Increased migratory and invasive ability is therefore an important characteristic for malignant cells and for their metastatic potential [7]. Metastasis is responsible for the majority of breast cancer patient deaths, rather than primary tumors, which in many cases can be surgically removed.The ability of tumor cells to generate metastases being an early event in the development of cancer, patients would benefit from treatments preventing or limiting the dissemination of malignant cells early on. Thus targeting the migratory phenotype is an interesting avenue for the improvement and development of novel tumor-suppressing treatments.

The migratory process relies on spatial and temporal signaling dynamics, for instance when interacting with the extracellular matrix (ECM) or signaling via growth factor receptors resulting in individual or collective cell movement [8], [9]. Gene expression studies have shown a regulatory response within 30 minutes of injury to murine or human tissue followed by hundreds of regulated genes a few hours thereafter [10], [11]. However, many essential events participating in cellular responses are a result of post-translational protein modification dynamics rather than changes in gene expression or protein abundance. In terms of the signaling mechanisms themselves, a phosphorylation response can occur within seconds of a perturbation [12], which can be studied using a time course data set. Several studies [13]–[15] have investigated changes in the phosphoproteome at early time points in response to cues such as growth factor stimulation, some as early as 5 seconds after perturbation.

Here we address the question of how cells sense injury from a signaling perspective and decide on directed cell migration in order to fill the void left following wounding. We propose to detect very early signaling dynamics following wounding, and aim to identify mechanisms involved in the earliest stages of the inflammatory response and leading to migration into the wound. As a model for response to wound healing and cell migration, we use an *in vitro* wound healing assay [16] applied to a confluent monolayer of the triple-negative breast cancer cell line MDA-MB-231 known to be invasive in

mouse models [17]. The scratch assay provides a model system to understand cell migration dynamics, also taking into account cell-cell interactions and their impact on overall migration into the gap. Currently, mechanistic insight regarding the way cells incorporate and process signals and how this is translated into cellular outcomes remains obscure. We sample the cellular phosphoproteomic networks for a series of time points, and we expect to help elucidate the overall signaling and system dynamics and provide a starting point for hypotheses requiring more in-depth targeted experimental design.

To study the signal transduction mechanisms involved using high- throughput global mass spectrometry, we enrich for tyrosine- as well as serine/threonine (Ser/Thr)-phosphorylated peptides from cells harvested at the earliest feasible time points and regularly up to one hour following scratching, totaling 10 time points. Given the small size of scratches (2x4 mm), we expected the fraction of the cell population sensing the perturbation to be relatively small, and therefore designed our study to maximize the phosphorylation signal related to wound sensing, while minimizing the response typically involved with sample handling-related stress [18]. Practically speaking, this was done by using a 384-well pin tool to reproducibly create 384 scratches in every tissue culture plate, thereby increasing the number of cells at the wound edge. Additionally, in order to capture these early cellular phosphorylation events, we snap froze cell monolayers in liquid nitrogen and immediately proceeded to lysis, thereby reducing confounding ischemia-driven signaling events.

We hypothesized that there exist several layers of complexity in the earliest response to scratching, and that cells experience these responses at different time points as a function of their distance to the wound. Our data display a very early upregulated phosphosite regulation occurring within minutes of wounding predominantly related to cell adhesion. Thereafter, a second large wave of activated signaling around 16-30 minutes occurs. We will present this dataset and point to the known cellular processes we have identified to be enriched within the very early response at the signaling level.

4.6 Results

4.6.1 Immediate phosphosite modulation upon wounding

As a model for response to wounding, we used the MDA-MB-231 breast cancer cell line and wounded the cell monolayer. Wounds closed 16 to 24 hours after scratching indicating that the closing of the wound is due to migration rather than proliferation (MDA-MB-231 cells have a doubling time of approximately 30 hours [19]). We therefore focused our study to the first hour post wounding in order to identify the signaling network changes which occur directly after wounding and that cause the initiation of the migratory response. In total, cells were harvested at 10 time points after wounding : immediately after scratching (denoted as "0", since this is the earliest feasible time point), 30 s, 45 s, 1 min, 2 min, 4 min, 8 min, 16 min, 30 min, and 1 h (Fig. 4.1). In order to preserve the phossphorylation state within cells [20] and to quench signaling at precise timepoints, we snap froze cells within seconds (approximately 5 s) following wounding. We believe this step to be critical, as most signaling related studies contain time-courses beginning after at least several minutes following perturbation, and are likely overlooking early signaling dynamics.

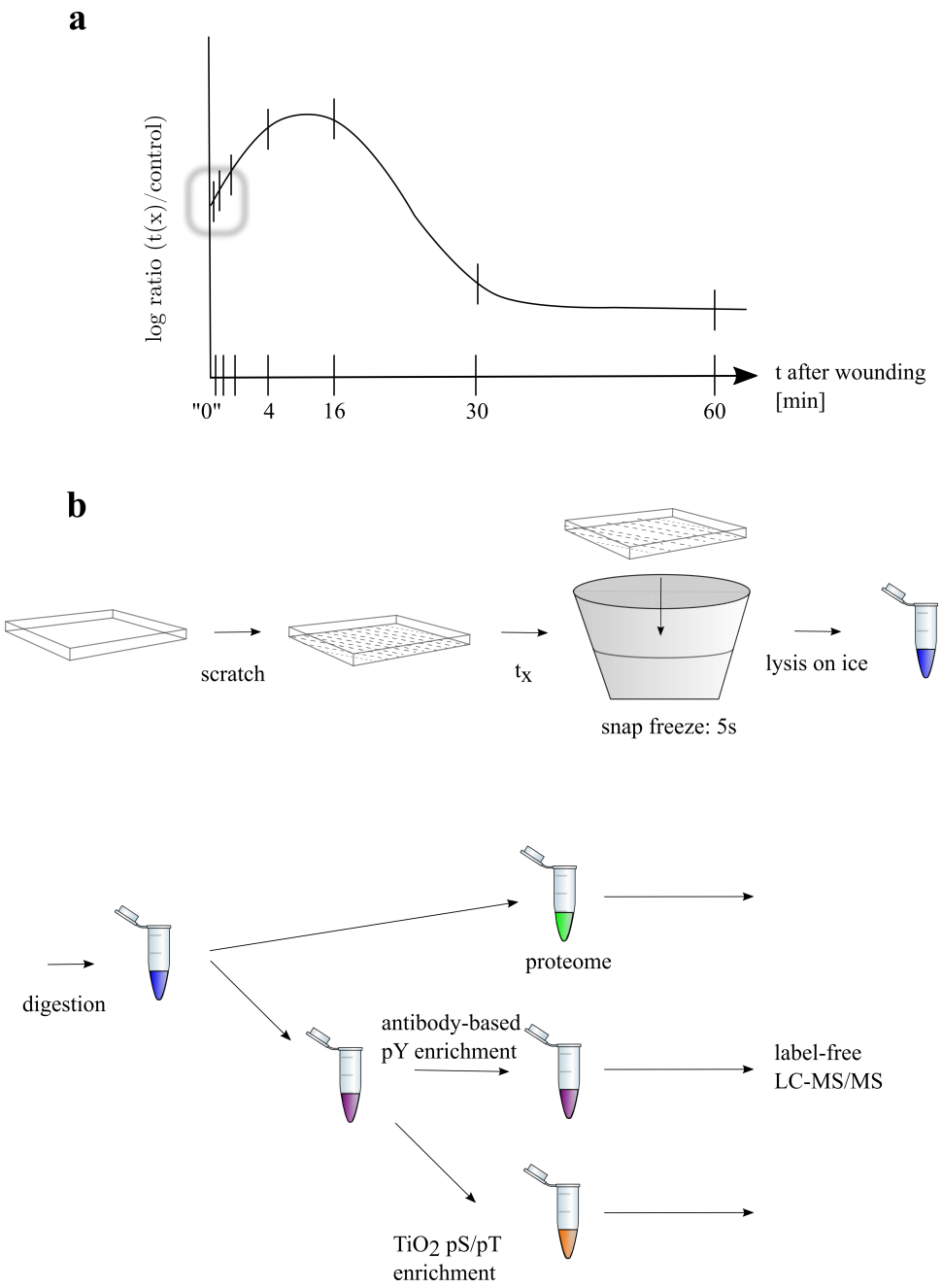


FIGURE 4.1: General illustration of the experimental workflow. **a** shows the full time-course for a putative phosphopeptide. The y-axis represents the log ratio of any given time point to the control : unwounded cells at 0 h. **b** depicts the lysis steps in detail as well as the subsequent proteomic workflow. Cell monolayers on tissue culture plates are immersed into liquid nitrogen for 3 s in order to retain the signaling state of cells.

Cells unwounded at 0 h served as controls, and the ratio of each time point to the control will be referred to as the label-free log fold change of a given peptide or protein throughout this study. After normalization of high confidence phosphopeptide ratios based on a linear model to account for small loading discrepancies, time-course data was fit using a Gaussian process to infer log ratios and the standard deviation for missing time points, which resulted in a total of 3333 phospho site temporal profiles, corresponding to 1768 proteins. Time-courses with significant variation with respect to the mean (variance > 0.005) were retained and denoted as being dynamically modulated. The sites with lower variation were classified as either "constantly modulated" if there was a significant fold change at all timepoints, or "unmodulated" in the case of temporal profiles that varied neither over time, nor compared to the non-wounded cells. The 1085 dynamically modulated temporal profiles were subjected to hierarchical clustering using a Euclidian distance metric 4.2. We observed a relatively small fold change compared to the control (Fig. 4.2), which may suggest that the wounding signal impacts a small number of cells most likely at the wound edge.

4.6.2 Dynamically modulated phosphoproteins exhibit distinct temporal activation profiles

The subset of peptides that were dynamically modulated could be further classified using hierarchical clustering to generate time-dependent signaling patterns based on the modulation of phosphopeptides normalized to their mean response over all tested timepoints. Seven clusters were derived by the dendrogram, and these were grouped into four temporal signaling trends, for which representative profiles are illustrated in the right panel of Fig. 4.2: "early" denotes 560 unique phosphopeptides with increased modulation at the earliest measurable time-point following wounding, and remained activated until approximately one to two minutes. "Mid" contains 85 unique phospho-peptides that displayed increased modulation at approximately two minutes, after which the signal returned to baseline. The "late" group contains 440 unique phosphorylated peptides showing significant modulation only at later timepoints, i.e. several minutes after wounding. Thus, over half of the varying sites were already modulated at the earliest measured time point. For simplicity, this time point was called "0", but in reality corresponds to approximately 5 seconds after wounding.

In over 24% of cases profiles of distinct phosphosites on the same protein fall in at least two different temporal clusters, thus displaying differential regulation, a phenomenon that has also been noted by others [21]. For example, in the case of EGFR, Tyr-1197 falls into the early cluster, Ser-1166 into the "mid" cluster, and Tyr-869 and Thr-693 show regulation in the latest time points. As observed in Olsen *et al.*, we detect tyrosine phosphorylation of EGFR to be upmodulated first, followed by activation of EGFR serine and threonine sites at later time points.

An additional example is MAPK1, a mediator of EGFR signaling, which was found to have modulated phosphopeptides belonging to different clusters. Phosphorylation at Tyr-187 is activated at "0" minutes, with a peak at 30 seconds before decreasing steadily. Thr-185 appears in the "mid" cluster, with a clear peak at 2 minutes, followed by a gradual decrease to slightly

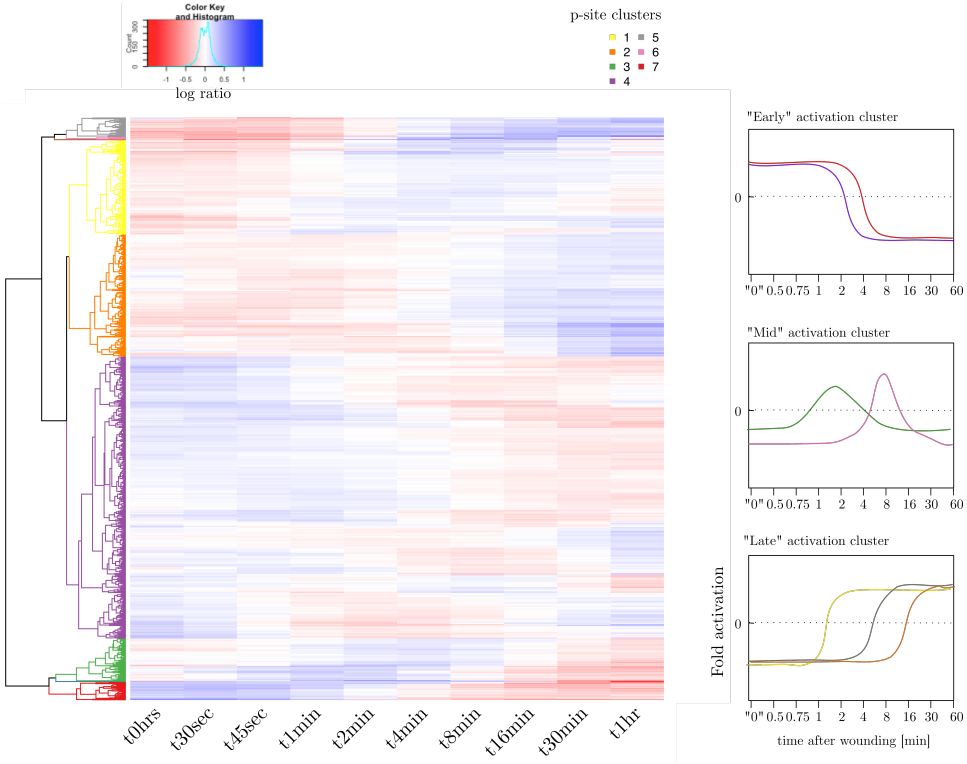


FIGURE 4.2: Heatmap representing the different signaling trends following cell monolayer wounding. 1085 phosphopeptides (vertical) whose ratio to the control (top left color key) showed variation over time were subjected to hierarchical clustering. Their corresponding temporal profiles are displayed horizontally. Clusters were generated by cutting the dendrogram at a height of 1.5, resulting in seven phosphosite subsets (denoted by colored dendrogram branches). These were then further classified into three groups representing early, mid, or late dynamics. The plots to the right of the heatmap show illustrative temporal profiles for the phosphosites in the corresponding "activation clusters". The dotted line corresponds to the non-varying phosphosites that were not used for hierarchical clustering.

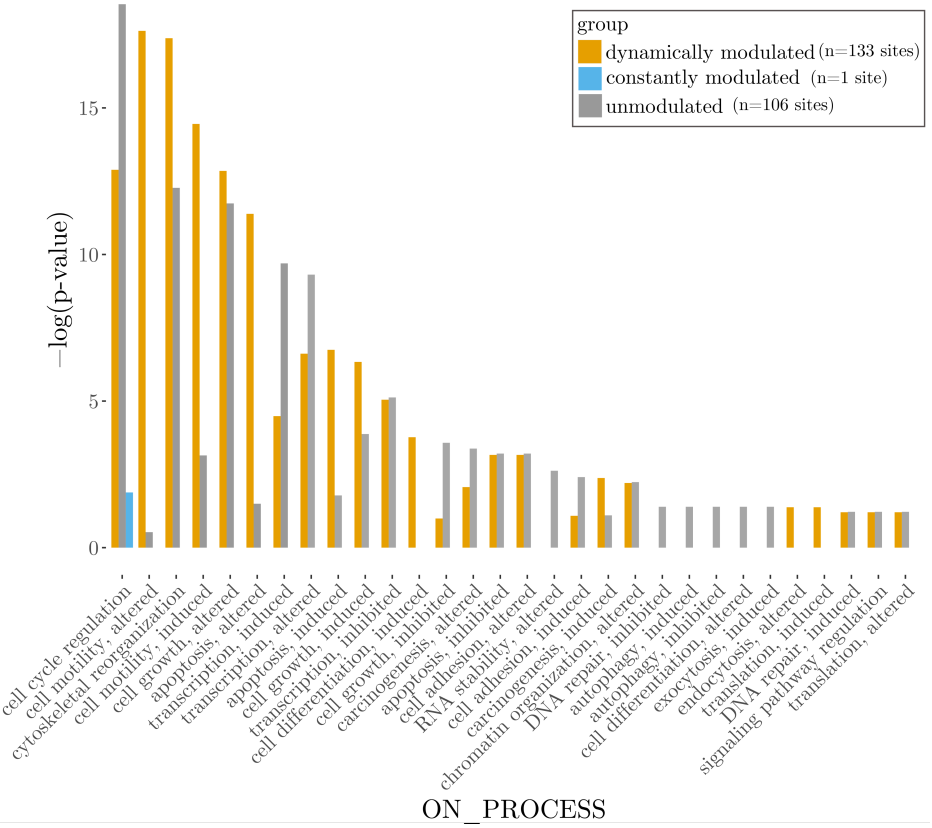


FIGURE 4.3: Regulatory information mapped from the PhosphoSitePlus database. Significantly enriched processes are shown for sites dynamically modulated (orange bars), sites constantly modulated (blue bars) as well as unmodulated sites showing low or no variation (grey bars). P-values were obtained by applying Fisher's exact test to sites mapping to processes in PhosphositePlus.

lower than pre-activation levels. Both sites are known to induce cell motility and growth.

4.6.3 Functional analysis of modulated phosphopeptides following wounding

Next we investigated the biological relevance of these phosphosites, by mapping to functionally annotated sites in the PhosphositePlus database [22]. Although over 77% of our phosphosites have been observed in the literature, the majority of the sites lack curated functional annotation, with only 3.3% of our phosphosites mapped to a known function.

Despite the low coverage of functionally annotated sites, we observed several significantly enriched processes as shown in Fig. 4.3. Strikingly, the top cellular process identified in the dynamically modulated phosphosites was induced cell motility. Indeed, the enrichment of phosphosites for increased cell motility was dramatic in comparison to non dynamically modulated sites. Cytoskeletal reorganization, apoptosis, cell growth and cell differentiation, were all also enriched in the dynamically modulated group.

4.6.4 Modulated phosphosites enrich for genes affecting wound closure

We further validated the functional relevance of phosphorylated proteins by investigating the effect of knocking down their corresponding genes using RNA interference on the velocity of cells after wounding. Interestingly only the modulated sites enriched for genes affecting cell velocity (p-value modulated sites: 0.017, p-value constantly modulated sites: 0.065, and p-value unmodulated sites: 0.703). Similarly phosphosites that were functionally annotated in PhosphositePlus as affecting cell motility or adhesion were also targeted with siRNA knockdown (Table 4.1). Of the 9 proteins that did not have modulated phosphosites, 33% significantly affected velocity post wounding. In contrast, of the 16 tested proteins that had dynamically modulated phosphosites 9 (56%) significantly affected cell velocity. Several of these function as signal transducers (EGFR, MAPK1, MAPK3, MPZL1, STK24, PRKCD) MAPK1 and MAPK3 are both modulators of EGFR signaling, and STK24 is a regulator of the JNK pathway. CD44 (related to invadopodia formation), CTNNA1 (component of adhesion complexes), and MYH9 (focal adhesion, cytoskeletal re-organization) are involved in the adhesive properties of cells.

4.6.5 Dynamically modulated phosphoproteins enrich for tyrosine kinases

Olsen *et al.* determined the proportion of tyrosine phosphorylation to make up approximately 1.8% of phosphorylated peptides. In our study peptides which did not change over time displayed a similar distribution of phosphorylated amino acids: 2.1% on tyrosine, 7.9% threonine and 90.0% serine, indeed close to previously reported tyrosine phosphorylation proportions. However, the phosphopeptides that were dynamically modulated showed an increase of phosphorylation on tyrosine residues, with 4.4% of sites being tyrosine, 11% threonine, and 84.8% serine residues. Table 4.2 lists the modulated tyrosine phosphosites and their corresponding gene names. Many of these tyrosine phosphorylated proteins are clearly migration-related, such as actin, Actinin alpha 1, Dynein light chain LC8-type 1 , Copine 1 and Plastin 3.

We further noted that proteins with tyrosine phosphorylation significantly affect velocity upon siRNA knock-down, underlining the importance of tyrosine phosphorylation in cellular communication and the migration processes. Of relevance additionally in the group are the tyrosine kinases which have in many cases been found to be overexpressed or mutated and lead to oncogenic states [23]. In particular we note EPHA2, which has been identified as a target for chemotherapy in the case of triple negative breast cancer. [24]. Finally, the tyrosine phosphorylation of p53 is noteworthy due to its established role as a tumor suppressor, but also due to its involvement in metastasis [25]. In addition to our p53 affecting cell velocity, exome sequencing of the MDA-MB-231 cells revealed the presence of the R280K oncogenic mutation in the p53 gene [26].

4.6.6 Functional analysis of proteins following wounding

As noted there are relatively few curated functional annotations in the Phosphosite Plus database. Gene Ontology (GO) biological process enrichment

Uniprot ID	Phospho-site	Gene	Process	Group	P-value	Delta velocity [px/h]
O00401	Y256	WASL (*)	cell adhesion, induced; cytoskeletal reorganization	unmodulated	3.75x10-4	4.82
Q9UPY8	S147, S162	MAPRE3 (*)	cell adhesion, induced; cytoskeletal reorganization	unmodulated	4.54x10-4	-2.92
O95297	Y262, Y263, Y96	MPZL1 (*)	cell adhesion, altered	unmodulated	4.76x10-4	-4.58
Q8IX03	S947	WWC1	cell motility, induced; cell growth, induced	unmodulated	1.13x10-2	2.85
Q9Y6Q9	S857	NCOA3	carcinogenesis, induced; cell motility, induced; transcription, induced	unmodulated	7.52x10-2	1.75
P23443	S424, S447	RPS6KB1	cell cycle regulation; cell adhesion, altered	unmodulated	1.53x10-1	-3.03
Q3V6T2	S1417	CCDC88A	cell motility, altered; cytoskeletal reorganization	unmodulated	2.14x10-1	1.60
P06748	T205, T208, T234, T237	NPM1	cell cycle regulation; cell motility, induced; cell growth, induced; cytoskeletal reorganization	unmodulated	7.34x10-1	-0.37
P60981	S3	DSTN	cytoskeletal reorganization; cell adhesion, altered	unmodulated	7.69x10-1	0.16
P28482	T185, Y187	MAPK1 (*)	cell motility, induced; transcription, induced; cell differentiation, induced; cell growth, induced; cell motility, altered; cytoskeletal reorganization; transcription, inhibited, cell growth, altered	modulated	1.68x10-7	-3.69
P00533	Y1197, Y869	EGFR (*)	cell motility, induced; cell growth, induced; cell growth, altered	modulated	1.09x10-6	6.01
Q05655	Y313	PRKCD (*)	apoptosis, altered; cell motility, induced; transcription, inhibited	modulated	7.40x10-6	-2.55

TABLE 4.1: Table of genes associated with phosphosites functionally annotated with migration in PhosphoSitePlus. Genes affecting velocity are marked with a star (*). Genes with a p-value lower than 0.005 are considered to affect the velocity of MDA-MB-231 cells upon knock-down. Delta velocity represents the change in migration speed [pixel/hour] compared to the siRNA non-targeting control.

Uniprot ID	Phospho-site	MS variance	Protein	Gene	Delta velocity [px/h]
Q8NFJS	317	0.2364	G protein-coupled receptor class C group 5 member A	GPRC5A	0.32
P28482	187	0.0270	Mitogen-activated protein kinase 1	MAPK1 (*)	-3.69
P50552	39	0.2199	Vasodilator-stimulated phosphoprotein	VASP	-0.49
P62191	25	0.0987	Proteasome 26S subunit ATPase 1	PSMC1 (*)	-7.54
P27361	204	0.0470	Mitogen-activated protein kinase 3	MAPK3	
P49023	118	0.2828	Paxillin	PXN	not tested
P49023	88	0.0964	Paxillin	PXN	not tested
P00533	1197	0.0075	**Epidermal growth factor receptor**	EGFR (*)	6.01
P00533	869	0.0205	**Epidermal growth factor receptor**	EGFR (*)	6.01
O43493	430	0.0064	Trans-golgi network protein 2	TGOLN2 (*)	-2.90
P30530	702	0.0254	**AXL receptor tyrosine kinase**	AXL (*)	-3.09
Q99829	391	0.0437	Copine 1	CPNE1	0.81
P53801	174	0.0677	Pituitary tumor-transforming 1 interacting protein	PTTG1IP	-1.42
P68104	141	0.0101	Eukaryotic translation elongation factor 1 alpha 1	EEF1A1 (*)	-5.44
P68104	29	0.0131	Eukaryotic translation elongation factor 1 alpha 1	EEF1A1 (*)	-5.44
Q03135	14	0.1398	Caveolin 1	CAV1	1.56
P07355	42	0.04944	Annexin A2	ANXA2	-0.25
P13797	127	0.0067	Plastin 3	PLS3	-0.55
Q6NZI2	308	0.0140	Polymerase I and transcript release factor	PTRF	0.16
Q8WWI1	170	0.0078	LIM domain 7	LMO7	-1.90
P60709	9	0.0248	Actin beta	ACTB	not tested
P29317	575	0.1454	**EPH receptor A2**	EPHA2	not tested
P29317	588	0.0132	**EPH receptor A2**	EPHA2	not tested
O95297	241	0.0169	Myelin protein zero like 1	MPZL1 (*)	-4.58
P18669	92	0.0200	Phosphoglycerate mutase 1	PGAM1	-3.29
P08238	56	0.0096	Heat shock protein 90 alpha family class B member 1	HSP90AB1	-0.99
Q09666	715	0.0205	AHNAK nucleoprotein	AHNAK	-2.44
P63167	65	0.0122	Dynein light chain LC8-type 1	DYNLL1	-0.41
Q96P48	231	0.0098	ArfGAP with RhoGAP domain ankyrin repeat and PH domain 1	ARAP1	-0.77
Q05655	313	0.0119	Protein kinase C delta	PRKCD (*)	-2.55
Q05655	334	0.0110	Protein kinase C delta	PRKCD (*)	-2.55
P06493	15	0.0261	Cyclin dependent kinase 1	CDK1 (*)	-7.67
P56945	387	0.0153	BCAR1 Cas family scaffolding protein	BCAR1	-5.73
P56945	410	0.0104	BCAR1 Cas family scaffolding protein	BCAR1	-5.73
P49840	279	0.0051	Glycogen synthase kinase 3 alpha	GSK3A (*)	3.03
Q96HC4	148	0.0059	PDZ and LIM domain 5	PDLIM5	-4.21
O00560	67	0.0095	Syndecan binding protein	SDCBP (*)	5.26
P07948	397	0.0100	**LYN proto-oncogene Src family tyrosine kinase**	LYN	2.46
Q05397	614	0.0666	Protein tyrosine kinase 2	PTK2 (*)	-6.77
Q05397	620	0.0205	Protein tyrosine kinase 2	PTK2 (*)	-6.77
Q13523	849	0.0427	Pre-mRNA processing factor 4B	PRPF4B (*)	-5.34
Q9UDY2	1149	0.0267	Tight junction protein 2	TJP2 (*)	-9.06
P12814	193	0.0621	Actinin alpha 1	ACTN1	-0.20
Q9NRY4	1105	0.0076	Rho GTPase activating protein 35	ARHGAP35	4.40
Q13813	2449	0.0071	Spectrin alpha non-erythrocytic 1	SPTAN1	-3.01
P04637	288	0.1638	Tumor protein p53	TP53 (*)	-2.95
P96QZ7	373	0.0084	Membrane associated guanylate kinase WW and PDZ domain containing 1	MAGI1	-3.22

TABLE 4.2: Table of tyrosine-phosphorylated sites dynamically modulated over time and their corresponding genes. Proteins in bold font are tyrosine kinases. Genes marked with a star were found to affect velocity of migrating cells upon knock-down. Delta velocity represents the change in migration speed [pixel/hour] compared to the siRNA non-targeting control.

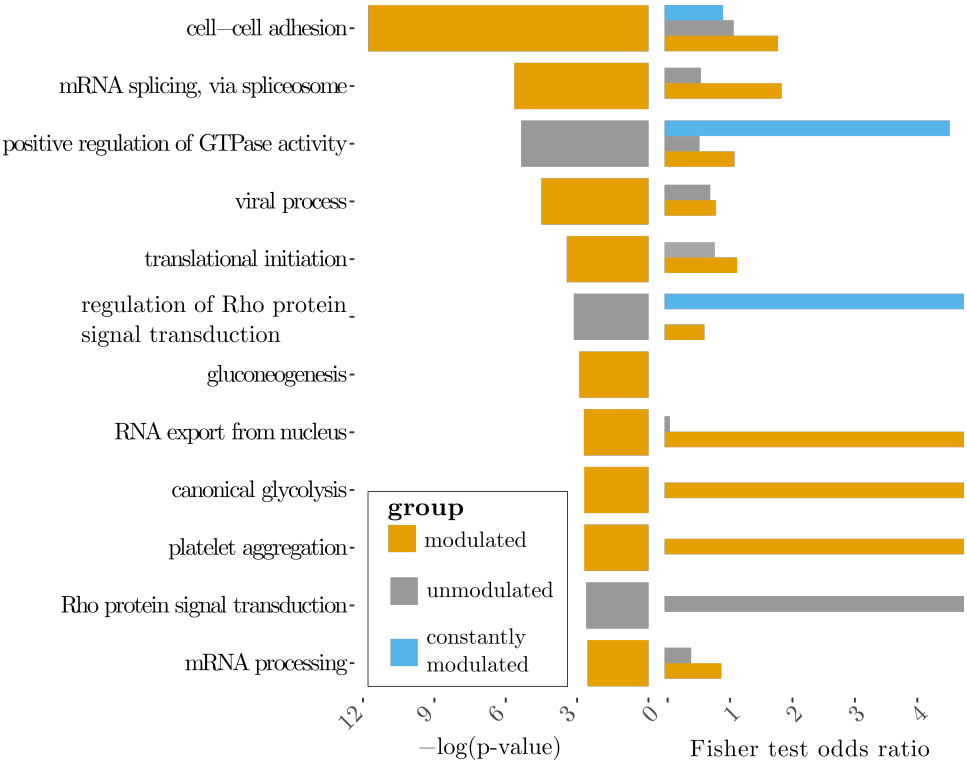


FIGURE 4.4: Left barplot: Gene Ontology Biological Processes significantly enriched in dynamically modulated sites (orange), and unmodulated sites (blue). No enriched processes were found to be significant in the group of constantly modulated phosphosites. Right barplot: For each process, the enrichment for number of genes significantly affecting velocity is shown. P-values were obtained from Gene Ontology enrichment analysis with a background genome comprised of genes detected in the unscratched cells.

analysis was therefore carried out on the genes corresponding to activated phosphosites involved in the various modulation clusters (Fig. 4.4). Cell-cell adhesion was the biological process with the largest activation in early time points and increased compared to the non-varying peptides, and thus confirmed that the phosphopeptides showing early signaling dynamics are indeed related to relevant processes as a result of the applied perturbation. Adhesion-related processes will be discussed further in subsection 4.6.7. Additional interesting processes were viral processes (potentially linked to an inflammatory response following perturbation), mRNA processing, splicing and export, translational initiation, as well as processes related to metabolism such as gluconeogenesis, and canonical glycolysis. For each process displaying enrichment (left barplot, Fig. 4.4), the enrichment for the number of genes significantly affecting velocity is shown (right barplot, Fig. 4.4). Dynamically modulated sites involved in cell-cell adhesion were twice as likely to affect velocity compared to sites in unmodulated or constantly modulated clusters. Genes confirmed to affect velocity are listed in Table 4.3.

4.6.7 Signaling Localization

The site modulation groups described above showed changes in the location of signaling activity as a function of their modulation after wounding (left barplot, Fig. 4.5). Supporting the previous results, the dynamic modulation response showed enrichment for clear migration-related cellular components such as cell-cell adherens junctions (p-value: 5.16x10-16), focal adhesion (p-value: 6.83x10-13), the extracellular matrix (p-value: 1.29x10-6), and cell-cell junctions (p-value: 1.84x10-3). First-line effectors of migration via their role in adhesion (cell-cell adherens junctions, focal adhesions, cell-cell junctions) showed an enrichment of genes affecting velocity. Additionally, components known to be involved in the migratory response were also found to be enriched in the dynamically modulated group, such as signaling in extracellular exosomes (p-value: 8.64x10-14), actin cytoskeleton (p-value: 1.17x10-6), the membrane and plasma membrane (respectively, p-value: 5.78x10-9 and p-value: 5.88x10-4), and the cell cortex (p-value: 1.16x10-4). As for the processes, the enrichment for the number of genes significantly affecting velocity is shown (right barplot, Fig. 4.5).

Interestingly, the unmodulated group of phosphosites displayed an enrichment in GTPase and Rho-related signaling when considering biological processes (Fig. 4.4). Diverse roles have been reported for the proteins of the Rho family, being at the same time responsible for cell motility via the creation of novel sites of adhesion as well as focal adhesions disassembly [9]. We noted that in Fig. 4.4, the processes related to GTPase and Rho signaling also displayed the most dramatic enrichment in genes affecting velocity upon knock-down. Furthermore, the categories for which this observation was also true in terms of cellular components (Fig. 4.5) was for the cell cortex and cytoskeleton. We therefore speculate that these might be linked. Velocity hits associated with significant GO processes are listed in Table 4.4.

Cell-cell adhesion was found to very strongly enriched for the dynamically modulated cluster genes (p-value: 1.67x10-12). In terms of cellular location, this would be expected to translate into signaling at cell-cell adherens junctions, which were also significantly enriched (p-value: 5.16x10-16).

We also observe regulation of phosphorylation on proteins neighboring cadherin complexes, such as tight junction proteins TJP1 and TJP2, which play a role in tight junctions as well as adherens junctions. Tight junctions provide a barrier function between cells and are placed apically in such a manner as to create tight seals between cells. Adherens junctions are placed just beneath the tight junctions, closer to the basal end of the cell. Together, these junctions form junctional complexes. Proteins such as TJP1 and TJP2 are thought to provide a scaffold for the placement of tight junctions with respect to other junctions [27].

We have shown that in our measurement of very early signaling we observed increased signaling at adherens junctions, as well as altered cell motility. One explanation for this could be that following wounding, the loss of neighboring cells with adhesions causes the cells to trigger migration in order to repair the wounded area and reconstitute the cell monolayer.

As seen in the processes in Fig. 4.3, the activated phosphosites were enriched for induction of cell motility compared to non-varying phosphosites. Acquisition of motility is known to be one of the first steps undertaken by cells in the process of migration and ultimately, of the genesis of metastases.

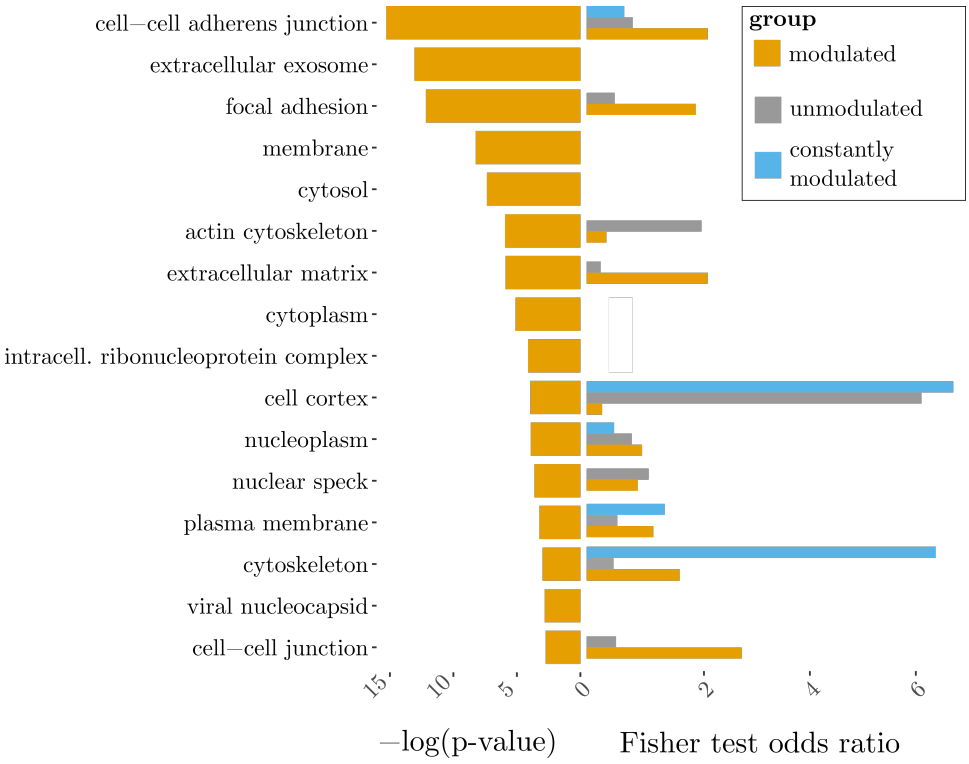


FIGURE 4.5: Gene Ontology Cellular Components significantly enriched in the dynamically modulated group (orange) and compared to unscratched cells. Processes were found to be enriched only in the group of dynamically modulated phosphosites. For each component, the enrichment for number of genes significantly affecting velocity is shown.

In order to achieve this goal, actin polymerization is required for cell polarization and elongation. In our data, signaling at the actin cytoskeleton (p-value: 1.17x10-6) was enriched in the modulated phosphosite group. The enrichment of cytoskeletal reorganization throughout the time-course (Fig. 4.3) could be due to the fact that elongation and assembly of actin is required in the early events of migration as well as during the next step : formation of focal adhesion. Then, focal adhesions (signaling enriched in modulated group with p-value: 6.83x10-13) are formed with the ECM (p-value: 1.29x10-6), as a requirement for contraction and movement of the main cell body [8]. Regulation of focal adhesion and cell migration can be mediated by different integrins depending on cell type, as well as other proteins such as CD44 or matrix metalloproteinases. Although we do not observe integrin regulation in this dataset, we do observe regulation of CD44 S706 in the early cluster, as well as a steady increase of non-modified CD44 over time. CD44 is known to mediate cell-cell and cell-matrix interactions via binding of hyaluronic acid, which in turn plays an important role in cell migration and promotion of metastasis, but not in the initiation of tumorigenesis [28]. Although focal adhesions are durable junctions, formed when there is ample time for strong attachment to a surface, our data does not suggest an increase in the formation of focal adhesions within minutes, but rather an increase in signaling detected at already existing focal adhesions at the sub-minute time-scale.

4.6.8 Extracellular matrix

The adherens and tight junctions described earlier mainly concern connections and communication between cells, allowing cells to transmit signal following a perturbation such as wounding. Another important event in relation to migration is the contact between cells and extracellular elements, typically mediated by integrins attached to actin filaments via talin proteins. In our data, we observe an initial dip of TLN1-S620 followed by steady increase over time. Moreover, in our data signaling modulated in the extracellular matrix was found to be rich in structural proteins (LMNA, MYH9, VIM, FLNA, FLNB).

Integrin signaling results in PTK2 activation, located downstream of several other growth factor receptors. PTK2, also known as focal adhesion kinase (FAK), can in turn regulate the AKT1 and the MAPK pathways, as well as promoting phosphorylation of SRC and BCAR1, among others. We observe regulation of PTK2 at several sites the dynamically modulated group. Moreover, phosphorylated BCAR1 sites show an increase in early time points. It should be noted that these events could also be the result of EGFR activation.

Interestingly, we also observed an enrichment in dynamic signaling in extracellular exosomes (p-value: 8.64x10-14), which allow for the transport of molecules from the cell to the extracellular space, and thus bilateral regulation of the cancer cell and the extracellular matrix surrounding it. We observed regulation of sites on kinases associated with extracellular exosomes, e.g. MAPK1, MAPK3, PRKAR2A, AXL, PRKCA, PRKCD.

Studies have identified an abundance of exosomes from cancer cells, and proteins contained in these exosomes can alter target cells or the microenvironment by transferring the oncogenic properties of the primary cell [29].

4.6.9 Discussion

To the best of our knowledge, this set of time points is the first and most extensive mass spectrometry-based phosphoproteomics dataset studying the early response to wounding. Previous studies investigating wound response have been based mainly on transcriptomics [10], [11], with time points starting several minutes after wounding. Olsen *et al.* described signaling dynamics downstream of EGFR using SILAC in combination with phosphopeptide enrichment. Indeed, the well-established SILAC workflow provides accurate quantification allowing precise determination of fold-changes upon perturbation. We show that reliable quantification of fold-changes upon perturbation is also feasible using label-free quantification in a phosphoproteomic dataset, which we have used to detect biologically relevant signaling trends.

Here we have demonstrated a strong enrichment for migration-associated phosphorylated proteins in dynamically regulated phosphosites and thus the validity of this dataset as a resource. However, many of the studied proteins that were identified as modulated at the level of their phosphorylation state are not recognized to have a role in relation to wound healing or migration. We believe that a number of the dynamically modulated sites lacking functional annotation at the level of their phosphorylation are also implicated in migratory or invasive processes with mechanisms remaining to be uncovered by further studies.

This dataset, although rich, suffers from an intrinsic property of most LC-MS/MS-based proteomics studies, in that the output is an average of a population of cells, and the spatial information we can gain is limited. Imaging methods (e.g. immunofluorescence, laser ablation coupled to mass cytometry) would provide the means to understand the directionality of the signal traveling away from the wound, as well as to observe protein dynamics within cells upon phosphorylation. This would allow further dissection of the signaling that appears to be unmodulated in this study: the total amount of the phosphorylated form of a given peptide may appear constant, while a change in e.g. cellular localization may increase its signaling impact. For further validation studies of this dataset, we will investigate the spatial dynamics of the modulated phosphorylation signal using imaging mass cytometry [30].

We recognize that due to the nature of the experimental setup the signaling presented here might be a distorted rendering of the normal physiological setting in the context of wounding. However, the very early modulation of signaling enriching for known migration-related processes and the confirmation thereof with our knock-down experiments in associated genes indicate that the regulation we observe is relevant to the study of migration of cancer cells. It should be noted that the type of cell motility assay may have an impact on the result and could thus be at least partly responsible for the importance of cell adhesion observed in this dataset. Indeed, cells are more dependent on adhesion in 2D assays such as the one used in this study compared to 3D cultures where cells are in contact with the extracellular matrix, thus providing other means of migrating.

In this work we demonstrate the feasibility of identifying novel phosphorylation sites in relation to the migratory response following wounding using high-throughput assays such as mass spectrometry. As touched upon earlier,

there is a paucity of phosphorylated residues with known functional annotation, limiting our prediction of mechanisms governing migration-related signaling. Utilizing antibody fragments specific to phosphorylation sites coupled with imaging akin to our knock-down assessing change in velocity, it would be possible to determine if these sites are affecting migration, and thereby work toward improving our functional knowledge of phosphorylation. As an illustration of this, we detect various phosphosites on EGFR to be modulated differently over the time course studied here. Impairing the function of each of these phosphosites separately or in combinations may lead to distinct phenotypic outcomes. Perhaps more importantly, these regulated sites may represent a pool of drug targets currently not taken advantage of. Indeed, most therapeutics aimed at kinases physically block the ATP binding site. Blocking specific sites of phosphorylation could be envisaged as an alternative means of intervention, for instance using site-specific antibodies. Given the centrality of signaling in migration, it is reasonable to ask whether targeted interference with phosphosites might be a strategy for the development of metastasis-decreasing therapeutics.

4.7 Author contributions

Conceptualization, F.V. and R.L.; Methodology, F.V., G.S., B.B., and J.L.; Software, M.E., T.A. and X.R.; Formal Analysis, M.E.; Investigation, F.V., G.S. and X.L.; Writing, Original Draft, F.V., J.L. and R.L.; Writing – Review & Editing, F.V., G.S., M.E., X.R., X.L., B.B., J.L. and R.L.; Supervision, J.L. and R.L.; Funding Acquisition, R.L.

4.8 Acknowledgements

The authors would like to thank Blagoy Blagoev for advice on the proteomic sample preparation protocol. This work was funded by the Innovation Fund Denmark Grand Solutions project; MorphoMap (1311-00010B). The authors report no conflicts of interest.

4.9 References

[1] H. Dvorak, "Tumors: Wounds That Do Not Heal", *The New England Journal of Medicine*, vol. 315, pp. 1650–1659, 1986.

[2] D. S. Dolberg, R. Hollingsworth, M. Hertle, and M. J. Bissell, "Wounding and its Role in RSV-Mediated Tumor Formation.", *Source: Science, New Series*, vol. 2303867116, no. 8, pp. 676–678, 1985.

[3] L. J. Dunham, "Cancer in Man at Site of Prior Benign Lesion of Skin or Mucous Membrane: A Review.", *Cancer Research*, no. 7, pp. 1359–74, 1972.

[4] K. E. De Visser, A. Eichten, and L. M. Coussens, "Paradoxical roles of the immune system during cancer development.", *Nature Reviews Cancer*, vol. 6, no. 1, pp. 24–37, 2006.

[5] D. Hanahan and R. A. Weinberg, "Hallmarks of Cancer: The Next Generation.", *Cell*, no. 5, pp. 646–74, 2011.

[6] S. Hussain, L. Hofseth, and C. Harris, "Radical Causes of Cancer.", *Nature Reviews Cancer*, vol. 3, pp. 276–85, 2003.

[7] Weigelt, Peterse, and van't Veer, "Breast cancer metastasis: markers and models.", *Nat Rev Caner*, vol. 5, no. 8, pp. 591–602, 2005.

[8] P. Friedl and K. Wolf, "Tumour-cell invasion and migration: diversity and escape mechanisms.", *Nature Reviews Cancer*, no. 5, pp. 362–74, 2003.

[9] J. Kassis, D. A. Lauffenburger, T. Turner, and A. Wells, "Tumor invasion as dysregulated cell motility.", *Seminars in Cancer Biology*, vol. 11, pp. 105–17, 2001.

[10] J. Cole, R. Tsou, K. Wallace, N. Gibran, and F. Isik, "Early gene expression profile of human skin to injury using high-density cDNA microarrays.", *Wound Repair and Regeneration*, 2001.

[11] L. Cooper, C. Johnson, F. Burslem, and P. Martin, "Wound healing and inflammation gene s revealed by array analysis of 'macrophage-less' PU.1 null mice.", *Genome Biology*, vol. 6, 2004.

[12] J. Dengjel, V. Akimov, B. Blagoev, and J. S. Andersen, "Signal transduction by growth factor receptors: Signaling in an instant.", *Cell Cycle*, no. 23, pp. 2913–6, 2007.

[13] B. Blagoev, S.-E. Ong, I. Kratchmarova, and M. Mann, "Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics.", *Nature Biotechnology*, vol. 22, no. 9, pp. 1139–45, 2004.

[14] Y. Zhang, A. Wolf-Yadlin, P. L. Ross, D. J. Pappin, J. Rush, D. A. Lauffenburger, and F. M. White, "Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules", *Molecular & Cellular Proteomics*, vol. 4, pp. 1240–1250, 2005.

[15] L. A. Chylek, V. Akimov, J. Rn Dengjel, K. T. G. Rigbolt, B. Hu, W. S. Hlavacek, and B. Blagoev, "Phosphorylation Site Dynamics of Early T-cell Receptor Signaling.", *PLoS ONE*, vol. 9, no. 8, 2014.

[16] C.-C. Liang, A. Y. Park, and J.-L. Guan, "In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro", *Nature Protocols*, 2007.

[17] K. M. Henkels, G. P. Boivin, E. S. Dudley, S. J. Berberich, and J. Gomez-Cambronero, "Phospholipase D (PLD) drives cell invasion, tumor growth and metastasis in a human breast cancer xenograph model", *Oncogene*, vol. 32, no. 4910, pp. 5551–5562, 2013.

[18] A. S. Gajadhar, H. Johnson, R. J. C. Slebos, K. Shaddox, K. Wiles, M. K. Washington, A. J. Herline, D. A. Levine, D. C. Liebler, and F. M. White, "Phosphotyrosine signaling analysis in human tumors is confounded by systemic ischemia-driven artifacts and intra-specimen heterogeneity.", *Cancer Research*, no. 7, pp. 1495–503, 2015.

[19] R. Limame, A. Wouters, B. Pauwels, E. Fransen, M. Peeters, F. Lardon, O. De Wever, P. Pauwels, and A. Ahmad, "Comparative Analysis of Dynamic Cell Viability, Migration and Invasion Assessments by Novel Real-Time Technology and Classic Endpoint Assays.", *PLoS ONE*, vol. 7, no. 10, e46536, 2012.

[20] C. Mueller, K. H. Edmiston, C. Carpenter, E. Gaffney, C. Ryan, R. Ward, S. White, L. Memeo, C. Colarossi, E. F. Petricoin, L. A. Liotta, and V. Espina, "One-Step Preservation of Phosphoproteins and Tissue Morphology at Room Temperature for Diagnostic and Research Specimens.", *PLoS One*, vol. 6, no. 8, e23780, 2011.

[21] J. V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, and M. Mann, "Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.", *Cell*, vol. 127, no. 3, pp. 635–48, Nov. 2006.

[22] P. V. Hornbeck, J. M. Kornhauser, S. Tkachev, B. Zhang, E. Skrzypek, B. Murray, V. Latham, and M. Sullivan, "PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.", *Nucleic Acids Research*, vol. 40, pp. D261–70, 2012.

[23] A. Gschwind, O. Fischer, and A. Ullrich, "The discovery of receptor tyrosine kinases: targets for cancer therapy.", *Nature Reviews Cancer*, vol. 4, no. 5, pp. 361–70, 2004.

[24] W. Song, Y. Hwang, V. Youngblood, R. Cook, J. Balko, J. Chen, and D. Brantley-Sieders, "Targeting EphA2 impairs cell cycle progression and growth of basal-like/triple-negative breast cancers.", *Oncogene*, 2017.

[25] S. Weissmueller, E. Manchado, M. Saborowski, J. P. Morris Iv, E. Wagenblast, C. A. Davis, S.-H. Moon, N. T. Pfister, D. F. Tschaharganeh, T. Kitzing, D. Aust, E. K. Markert, J. Wu, S. M. Grimmond, C. Pilarsky, C. Prives, A. V. Biankin, S. W. Lowe, and H. Hughes, "Mutant p53 drives pancreatic cancer metastasis through cell- autonomous PDGF receptor beta signaling.", *Cell*, vol. 157, no. 2, pp. 382–394, 2014.

[26] L. Hui, Y. Zheng, Y. Yan, J. Bargonetti, and D. Foster, "Mutant p53 in MDA-MB-231 breast cancer cells is stabilized by elevated phospholipase D activity and contributes to survival signals generated by phospholipase D", *Oncogene*, vol. 25, pp. 7305–7310, 2006.

[27] A. Hartsock and W. J. Nelson, "Adherens and tight junctions: Structure, function and connections to the actin cytoskeleton.", *Biochimica et Biophysica Acta - Biomembranes*, no. 3, pp. 660–9, 2008.

[28] H. Ponta, L. Sherman, and P. A. Herrlich, "CD44: From adhesion molecules to signalling regulators.", *Nature Reviews Molecular Cell Biology*, no. 1, pp. 33–45, 2003.

[29] Y. Soung, T. Nguyen, H. Cao, J. Lee, and Chung J, "Emerging roles of exosomes in cancer invasion and metastasis.", *BMB Reports*, vol. 49, no. 1, pp. 18–25, 2016.

[30] C. Giesen, H. Wang, D. Schapiro, N. Zivanovic, A. Jacobs, B. Hattendorf, P. Schuffler, D. Grolimund, J. Buhmann, S. Brandt, Z. Varga, P. Wild, D. Gunther, and B. Bodenmiller, "Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry", *Nature Methods*, vol. 11, no. 4, pp. 417–22, 2014.

[31] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: A flexible trimmer for Illumina sequence data.", *Bioinformatics*, no. 15, pp. 2114–20, 2014.

[32] J. Li, U. Rix, B. Fang, Y. Bai, A. Edwards, J. Colinge, K. L. Bennett, J. Gao, L. Song, S. Eschrich, G. Superti-Furga, J. Koomen, and E. B. Haura, "A chemical and phosphoproteomic characterization of dasatinib action in lung cancer.", *Nature Chemical Biology*, vol. 6, no. 4, pp. 291–9, 2010.

[33] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo, "The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.", *Genome Research*, no. 9, pp. 1297–303, 2010.

[34] M. A. Depristo, E. Banks, R. E. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, A. Mckenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly, "A framework for variation discovery and genotyping using next-generation DNA sequencing data.", *Nature Genetics*, no. 5, pp. 491–8, 2011.

[35] W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, and F. Cunningham, "The Ensembl Variant Effect Predictor.", *Genome Biology*, no. 1, p. 122, 2016.

[36] S. Tyanova, T. Temu, and J. Cox, "The MaxQuant computational platform for mass spectrometrybased shotgun proteomics.", *Nature Protocols*, vol. 11, no. 12, pp. 2301–2319, 2016.

[37] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016.

[38] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik, *Cluster: Cluster analysis basics and extensions*, 2016.

[39] S. Durinck, P. Spellman, E. Birney, and W. Huber, "Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt.", *Nature Protocols*, pp. 1184–1191, 2009.

[40] Wickham and Hadley, "Reshaping data with the reshape package", *Journal of Statistical Software*, vol. 21, no. 12, 2007.

[41] ——, "Ggplot2: Elegant graphics for data analysis", 2009.

[42] H. Wickham and R. Francois, *Dplyr: A grammar of data manipulation*, R package version 0.5.0, 2016.

[43] C. Rasmussen and C. Williams, *Gaussian processes for machine learning*. MIT Press, 2006.

[44] C. Zhu, R. Byrd, and J. Nocedal, "L-BFGS-B: Algorithm 778: L-BFGS-B FORTRAN routines for large scale bound constrained optimization", *ACM Transactions on Mathematical Software*, no. 4, pp. 550–560, 1997.

4.10 Materials and Methods

4.10.1 Cell culture

MDA-MB-231 cells were grown in DMEM-F12 (1:1) media for SILAC (Thermo #88215) supplemented with Pen/Strep (Gibco, #15140-122) and 10% dialysed fetal bovine serum (Sigma #F0392). In preparation for the wound assay, a constant number of cells were plated in rectangular tissue culture dishes (Nunc, 165218) in order to create reproducible scratch patterns. Cells were allowed to reach close to confluency for 48 hours, were serum-starved for 24 hours, after which serum was re-introduced. 24 hours after ending serum starvation, the cell monolayer was scratched using either a 96-pin (late time points: t=8min and upwards) or 384-pin tool (early time points: t=0 until 4min) (V&P, 96 and 384 wounding tools, with pins #FP3-WP). When using the 96-pin tool, 4 scratches (2mm X 4mm) were made in close succession amounting to a total of 384 scratches. Reproducible scratching was ensured by using a pin guide tailored to the culture dishes (designed & produced in-house).

4.10.2 Sample preparation

Immediately after scratching, media was replaced with fresh serum-free media (if early) or with serum-containing media (later time points). After the desired time elapsed, cells were washed with serum-free media, and placed in liquid nitrogen in order to rapidly quench cellular activity. On ice, cells were scraped off plates into a modified RIPA buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 1 mM EDTA) supplemented with phosphatase inhibitors (5 mM β-glycerophosphate, 5 mM NaF, 1 mM Na3VO4) and protease inhibitors (Roche cOmplete ULTRA Tablets, EDTA-free, #05892791001), sonicated, and spun down for 20 min at 4300G. The time series samples were processed on the same day and replicates were created weeks apart repeating the same procedure, due to the amount of protein lysate required for proteomic analyses. This likely resulted in a decreased correlation between replicates. Ice-cold acetone was added to the supernatant to achieve a final concentration of 80% acetone, and protein was left to precipitate overnight at -20°. Precipitated protein was pelleted by centrifugation at 1800G for 5 min and solubilized in 6 M urea, 2 M thiourea, 10 mM HEPES pH 8.0. Protein was quantified using the Bradford assay and 24 mg of each sample were reduced with 1 mM DTT, alkylated with 5 mM ClAA, and digested with endopeptidase Lys-C (1:200 v/v, Wako #129-02541) for 3 hours. Samples were diluted to 1 mg/mL protein using 50 mM ammonium bicarbonate and incubated overnight with trypsin (1:200 v/v, Sigma #T6567). Digested samples were acidified and urea removed using SepPak (Waters #WAT020515) C18 cartridges. Peptides were eluted, an aliquot of 100μg set aside for proteome analysis, and the remainder dried by lyophilization, before

resolubilization in commercial IAP buffer for phospho-tyrosine peptide enrichment, following the manufacturer's indications (Cell Signaling, #8803). Phospho-tyrosine enriched peptides were eluted and cleaned up using C18 StageTips (3M Empore, #66883-U).

Unbound peptides from the immunoprecipitation were retained and diluted ten-fold before being put through SepPak C18 cartridges. Eluates were made up to 80% ACN, 6% TFA and TiO_2 beads (GL Sciences, #5020-75010) in loading buffer (0.02 g/mL Dihydroxybenzoic acid, 30% CH3CN, 4% TFA (or: 0.13 M Dihydroxybenzoic acid, 30% CH3CN, 4% TFA)) were added at a 1:1 w/w ratio with peptides, and the peptide-bead slurry was rotated end-over-end at RT for 30 minutes. Unbound peptides were collected for a second and third enrichment using TiO_2 microspheres. Beads were washed three times with 60% ACN, 1% TFA, transferred to a C8 96-well plate (3M Empore, #6314), and phosphopeptides were eluted sequentially with two elution buffers: 5% NH_4OH, followed by 10% NH_4OH, 25% ACN. Eluates were combined and cleaned up using C18 StageTips.

Peptides from proteome and phosphoproteome samples were quantified using the Pierce quantitative colorimetric peptide assay in order to standardize final loading amounts. Samples were lyophilized and equalized in 1% formic acid (FA) to same loading amounts for label-free quantification. Immediately prior to injection, spike-in peptide standards (Sigma, MSP1H) were added to each sample for normalization purposes with a final on-column load of 33 fmol. Due to the lower amount of peptides expected following phosphotyrosine enrichment, these samples were not quantified but their volumes were equalized prior to MS analysis. Due to the nature of label-free quantification, standardized conditions were achieved by processing samples in parallel, as well as monitoring of chromatographic and mass spectrometric conditions over time using a complex digest of HeLa cells.

4.10.3 MS data acquisition

All spectra were acquired on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) operated in data-dependent mode coupled to an EASY-nLC 1200 (Thermo Fisher Scientific) liquid chromatography pump and separated on a 50cm reversed phase column (Thermo, PepMap RSLC C18, 2μM, 100Å, 75μmx50cm). Proteome samples (non-enriched) were eluted over a linear gradient ranging from 0-11% ACN over 70 min, 11-20% ACN for 80 min, 21-30% ACN for 50 min, 31-48% ACN for 30 min, followed by 76% ACN for the final 10 min with a flow rate of 250 nl/min. Phosphopeptide- or phosphotyrosine enriched samples were eluted over a linear gradient ranging from 0-18% ACN over 195 min, 18-26% ACN for 30min, 26-76% ACN for 10min, followed by 76% ACN for the final 5 min with a flow rate of 250 nL/min.

Survey full scan MS spectra were acquired in the Orbitrap at a resolution of 120000 from m/z 350-2000, AGC target of 4e5 ions, and maximum injection time of 20 ms. Precursors were filtered based on charge state (≥ 2) and monoisotopic peak assignment, and dynamic exclusion was applied for 60s. A decision tree method allowed fragmentation for ITMS2 via ETD or HCD, depending on charge state and m/z. Precursor ions were isolated with the quadrupole set to an isolation width of 1.6 m/z. MS2 spectra fragmented by ETD and HCD (35% collision energy) were acquired in the ion trap with an

AGC target of 1e4. Maximum injection time for HCD and ETD was 200 ms for enriched samples, and 80ms for proteome samples.

4.10.4 Fasta file generation

DNA was extracted from the MDA-MB 231 cell line at 80% confluence using the Qiagen QIAamp DNA Mini kit, according to the manufacturers' instructions. Sequencing was performed by Beckman Coulter Genomics with a target average exome coverage of 89x. We filtered the reads with Trimmomatic [31] with the following parameters: headcrop = 3, minlen = 30, trailing = 3. We aligned the trimmed reads to the hg19 reference genome with BWA [32], applied GATK [33] base quality score recalibration, indel realignment, duplicate removal, and performed SNP and INDEL discovery and genotyping across all samples simultaneously using standard hard filtering parameters according to GATK Best Practices recommendations [34]. We used the Ensembl Variant Effect Predictor (VEP) [35] with Ensembl v. 75 to predict the effect of the mutations on the protein sequence. In order to find non-reference, mutated peptides in the MS data, we increased the search Fasta file with mutations affecting the protein sequence detected by WES with a high sensitivity filter: QD $\leq$ 1.5, FS $\geq$ 60, MQ $\geq$ 40, MQRankSum $\leq$ -12.5, ReadPosRankSum $\leq$ -8.0, and DP at least 5 / sample on average.

4.10.5 Mass spectrometry identification and quantification

All label-free raw data were searched using MaxQuant v1.5.1.2 [36] against the human Ensembl database (v88), as well as the MDA-MB-231 specific proteins from exome sequencing. Default settings were used with the exception of a minimum peptide length of 6 amino acids. STY phosphorylation was added as a variable modification. An FDR of 0.01 was applied at the level of proteins, peptides and modifications. Phosphosites were filtered for a localization score of at least 0.9. Data processing was carried out in R, with additional packages [37]–[42].

4.10.6 Markov Chain Monte Carlo

A likelihood function for each concentration c or phosphorylation occupancy o ratio was calculated as follows:

$$
\begin{aligned}
L(c, \bar{o}, \bar{o}') &= \log P(\bar{z} | \bar{\mu}(c, \bar{o}, \bar{o}')) \\
&= \sum_{i=1}^{I} \Bigg(\tilde{\nu}_i \log \Bigg[1 + \gamma_i \Big(\bar{x}_i - c - \\
&\qquad \sum_s I_{is} \Bigg[\log \left(\frac{1 - o'_s}{1 - o_s} \right) + t_{is} \left(\log \left(\frac{o'_s}{1 - o'_s} \right) - \log \left(\frac{o_s}{1 - o_s} \right) \right) \Bigg] \Big)^2 \Bigg] + \\
&\qquad \frac{1}{2} \log(\gamma_i) \Bigg)
\end{aligned}
\tag{4.1}
$$

where $\hat{x}$ is a vector of the peptide ratios mapped to the protein, I and t are indicator matrices of size $i \times s$ reporting whether peptide i spans phosphosite

$s\,(I)$ or is modified on site $s\,(t)$, and ν and γ are Bayesian variance parameters established by fitting Beta distributions by binned absolute log ratios. We sampled this likelihood for each protein separately with Bayesian Markov Chain Monte Carlo (MCMC) with Jeffrey priors on o and o' with $\alpha_1 = \alpha_2 = \frac{1}{2}$, and an exponential prior with $\lambda = 2$ on c.

Finally we calculated the means and standard deviations of the chains, resulting in means and standard errors for each protein and phospho-site occupancy. We calculated phosphosite ratios for a given sample to the control as $log(o.sample/(1 - o.sample)) - log(o.reference/(1 - o.reference))$.

4.10.7 Gaussian Process Regression

Individual protein concentrations and phosphorylation occupancies from Markov Chain Monte Carlo peptide integration were calculated for each replica set of MS experiments. The time point values X (concentration and occupancy) were considered gaussian random variables. For each protein and phospho-site at separate time points the mean value and standard error of the MCMC chains after burn-in constituted the data expectation value $\overline{X}$ and variance ${\sigma_X}^2$.

Given the non-linear time intervals in the MS data a new unit time was assigned to the ordered time points. Before modeling, the time series were normalized by subtracting the series mean value $\overline{X(t)}$. Only series with more than one time point were kept for modeling.

Gaussian Process Regression (GPR) [43] was used to model each time series as a function of unit time given the observed data

$$P(t) = \mathrm{GPR}(t|X(t), \sigma_{X(t)}).$$

This Bayesian approach allowed modeling the data and noise collectively, and enabled posterior predictions via the chosen prior kernel. The kernel hyper-parameters (general noise level and length scale) were optimized with respect to the GPR log-marginal likelihood using the L-BFGS-B algorithm [44]. The best model of three random initialized runs within the parameter constraints was chosen. Time series imputation of all missing data points was performed with the optimized GPR function $P(t)$ to produce complete time series including noise estimates.

4.10.8 Prior Kernel

The prior used was a sum of a squared exponential kernel [43] and a general noise component as a ('white kernel')

$$k(x_i, x_j) = \exp\left(-\frac{1}{2}d(x_i/l, x_j/l)^2\right) + k(\mathrm{noise}|x_i = x_j).$$

where d is the euclidian distance and l is the length-scale parameter for the exponential part of the kernel, and reports on the similarity between consecutive points. It must be positive ($l > 0$) and is fitted during optimization.

The white kernel was constrained to the range $[min({\sigma_{X(x)}}^2), max({\sigma_{X(x)}}^2)]$. The prior covariance matrix was initialized at the diagonal with the data noise ${\sigma_{X(x)}}^2$. The squared exponential kernel length scale was constrained to $[1, 2]$ time units.

4.10.9 Functional enrichment analyses

Gene Ontology categories with an FDR <5 were considered, and p-value $<$ 0.05 was used as significance measure. Calculation of enrichment for processes in PhosphositePlus was done using Fisher's exact test and p-values $<$ 0.05 were considered significant. Calculation of enrichment of genes affecting velocity was done using Fisher's exact test and categories with an odds ratio $>$ 1 were deemed significant. All analyses were carried out using the R programming language[37].

4.10.10 Tissue culture of fluorescent cells

MDA-MB-231 Cells expressing cytoplasmic green fluorescent protein and red fluorescent protein-tagged Histone H2 were cultured in T-75 flasks and passaged twice a week. Passaging was performed by firstly removing the medium and washing the cell sheet with phosphate buffered saline (PBS, Thermo Fisher Scientific). Cells were then incubated with 2ml of Trypsin (Thermo Fisher Scientific) for 1-2 minutes at 37°. Flasks were struck in order to dislodge the cell sheet prior to the addition of 3ml of medium (DMEM/F12, Life Technologies) to inactivate the Trypsin. Cells were transferred to a new flask and fresh medium added to give a final volume of 10ml.

4.10.11 RNAi screening

Silencer Select siRNAs (Thermo Fisher Scientific) were prepared by combining three individual siRNAs for each gene into an equimolar pool. Pooled siRNAs were diluted in OptiMEM medium (Thermo Fisher Scientific) before being combined 1:1 with transfection reagent (Lipofectamine RNAiMAX, Thermo Fisher Scientific). The siRNA/reagent transfection mixture was incubated at room temperature for 15 minutes before 5μL of mixture were added to 384-well CellCarrier plates (PerkinElmer). Cells were diluted to the required density in medium and 35μL added to each well to give a final concentration of 10nM siRNA and 0.06μl of transfection reagent. The cells were incubated with siRNAs for 48 hours, at 37°C, 5% CO2, 95% humidity prior to wounding and analysis of cell velocity.

4.10.12 Cell velocity

Wound healing assays were performed by scratching the cells with a 96-well pin tool to create wounds 2mm in thickness. Cells were imaged on the Opera (PerkinElmer), at 37°C, 5% CO2, immediately after wounding and every hour for 16 hours using a 10X objective with 2-pixel camera binning and 3 images captured per well. A 488nm excitation line with 450/50nm emission filter was used for green fluorescent protein detection and a 561nm excitation line with 585/40nm emission filter for red fluorescent protein detection. Analysis of cell velocity was performed using the Bowhead algorithm (bowhead.lindinglab.org). A p-value $<$ 0.005 was used to delineate genes affecting velocity of wound closure.

GO Biological Process	Group	Genes
Cell-cell adhesion	Modulated	MLLT4, KRT18, PAK4, RPL14, DBN1, EIF5, KTN1, STK24, RTN4, SDCBP, ERC1, TJP2, HCFC1, LARP1, MPRIP, PKM2, ATXN2L, PPP1R13L
Positive regulation of GTPase activity	Modulated	MLLT4, PTK2, ARHGAP12, RAB3GAP1, SHC1, ARHGEF18, EGFR
mRNA splicing, via spliceosome	Modulated	RAVER1, HNRNPL, SNRNP200, RBM15, SF3B1, U2AF2, DBR1, HNRNPA2B1, ARS2, HNRNPUL1, SRRM2, PRPF4B, PCBP2, SRSF1
Positive regulation of GTPase activity	Constantly modulated	GAPVD1, ARHGEF18, FNBP1
Regulation of Rho protein signal transduction	Constantly modulated	ARHGEF18
Cell-cell adhesion	Unmodulated	MLLT4, CALD1, PAK4, DBN1, EIF5, KTN1, CDC42EP1, STK24, ZC3HAV1, TJP2, PKN2, LARP1, MPRIP, IQGAP1, ATXN2L, PPP1R13L
Rho protein signal transduction	Unmodulated	ROCK1, BCL6, CDC42EP1, ROCK2

TABLE 4.3: Phosphopeptide regulation categories from the GO biological processes Fig. 4.4 with a Fisher's test odds ratio >1 for the enrichment genes affecting velocity were selected and displayed. Genes may belong to multiple categories due to the phosphosites they represent.

GO Cellular Component	Group	Genes
Cell-cell adherens junction	Modulated	MLLT4, KRT18, PAK4, RPL14, DBN1, EIF5, CTNNA1, KTN1, STK24, RTN4, SDCBP, MYH9, ERC1, TJP2, HCFC1, LARP1, MPRIP, PKM2, PKP2, EGFR, ATXN2L, PPP1R13L, TLN1
Cell-cell junction	Modulated	MLLT4, SH3KBP1, PRKCD, TLN1, PKP2, CTNNA1
Cytoskeleton	Modulated	ARHGEF18, AKAP12, PTK2, MAPK1, KIAA0802, SH3KBP1, HSPB1, SYMPK, SMTN, SDCBP
Extracellular matrix	Modulated	RPS3A, CANX, RPS3, HSPB1, RPL12, RPS25, RPL30, CCT2, LGALS1, CCT6A, MYH9, PKM2, PRKDC, RPS5
Focal adhesion	Modulated	PAK4, PTK2, RPL30, CLASP2, CTNNA1, RPS3A, HSPB1, RPS3, PRKAR2A, SDCBP, MYH9, CLASP1, RPL4, PPFIA1, AKAP12, MPRIP, MPZL1, RPS5, MAPK1, RPL18, EGFR, CD44, SH3KBP1, RPL12, PCBP2, HMGA1, TLN1

TABLE 4.4: Phosphopeptide regulation categories from the GO cellular components Fig. 4.5 with a Fisher's test odds ratio >1 for the enrichment genes affecting velocity were selected and displayed. Genes may belong to multiple categories due to the phosphosites they represent.

Part III

Epilogue

Chapter 5

Summary and Perspectives

In this thesis, I have focused on two main projects utilizing mass spectrometry-based proteomics to measure changes at the protein and phosphorylated protein levels. The first piece of work laid out in Chapter 3 looked at changes upon expression of the PRKAR1A-RET fusion found in thyroid cancer. I showed that the expression of this chromosomal rearrangement at endogenous levels caused immortalized cells to acquire characteristics of transformed cells. I was able to identify and propose several mechanisms by which the protein-level effects of this major genetic event might push cells toward transformation and thus a state of increased malignancy.

The elucidation of the signaling network changes driving a more cancerous phenotype will be key to better understand the functional effects of this relativelly understudies class of genetic alterations.While this work is only a first step in terms of understanding the system-wide repercussions of a single gene fusion, it is one of the few studies that have used high-throughput proteomics to investigate its effects. Given the enrichment of kinases as fusion partners with oncogenic behavior, in the future it will be crucial to study a much larger number of chromosomal rearrangements in order to uncover trends in their modulation of protein networks and the cancer hallmarks they thereby affect.

In the second manuscript (Chapter 4) I induced a migratory signaling response following wounding of triple negative breast cancer cells, and validated this both *in silico* and experimentally. I also identified numerous regulated protein phosphorylation sites that have not yet been linked to migration. Follow-up experiments are already in place for the validation of a subset of these sites, as well as to gain a deeper understanding of the spatial dynamics of very early signaling upon wounding. In terms of the application of these findings in the clinic, it is interesting that there are currently no available treatments targeting specific phosphorylation sites. This is likely due to the difficulty in developing drugs targeted to one site. We propose to develop antibody fragments toward specific phosphorylation sites and to then inhibit a site alone or in conjunction with other sites. It is possible that this could be sufficient to affect signal transduction and therefore to halt disease initiation or progression. This strategy would differentiate itself from kinase inhibitors, which generally target the active site of kinases. During this work, it became very apparent that the number of functionally annotated sites is limited. Moving forward it will be essential that studies investigating signaling link sites to biological functions, rather than simply listing these if they are to be a resource to the community. For instance, in the work performed in this thesis, we suggest that the modulated phosphosites have a role in the induction of migration.

The work described herein underscores the importance of studying systems from a protein perspective within the field of cancer research. Indeed, as alluded to throughout the thesis, there is a discrepancy between the approach of determining drug targets and the way in which drugs function *in vivo*. For the most part, anti-cancer drugs typically exert their effect by interacting with a protein, thereby affecting the dysregulated behaviour of the malignant cell. Interfering with processes by targeting proteins is a powerful approach, given that proteins are the effectors of cell behavior and directly affect the very phenotypes that are to be targeted. Is it curious then that drug targets are often selected based on findings from genomic or transcriptomic data. While this makes sense from a historical perspective, it is perhaps now time to turn the tables.