# ABSTRACT


Title of dissertation: ANALYZING BIOLOGICAL NOISE
Ashutosh Gupta, Doctor of Philosophy, 2012

Dissertation directed by: Dr David L Levens
National Cancer Institute, NIH, Bethesda

Dr Arpita Upadhyaya
University of Maryland, College Park



Biological systems are remarkably precise in a lot of different ways. Not only do organisms have the capacity to reproduce, they also have the capacity to defend themselves from external factors. The capability to fight diseases, in particular the immune system, is an integral part of the evolution and natural selection in all plants and animals. For most species there are multiple layers of defense, which are adaptive and provide mechanisms (or adaptive immunological memory) to remember previous attacks and successively improve the response. From reproduction to defense and maintenance, each organism constantly monitors its internal and external environments at several different levels. Several crucial constituent factors are required to be maintained at close tolerances. A deviation, or a push, away from equilibrium could prove fatal to an individual cell or the whole organism. These deviations also have a shared history with our evolution in the form of diseases like *cancer*.

In this study, we present some of our efforts to understand the origin and

control of this biological noise at four different levels from a physical sciences perspective.

The entire study of this dissertation has its origins linked to a proto-oncogene called *c-myc*, which is believed to regulate about 10% of mammalian genes. It controls all major decisions of cells, including cell division and cell death, and it is known to be deregulated in most types of cancers. Noisy *c-myc* transcription can have disastrous effects, thus its expression levels must be controlled very tightly by cells.

At the DNA level, we examine a dynamic feedback mechanism where DNA supercoils during transcription, and dynamic torsional stresses are mechanically coupled with ongoing transcription to control the transcriptional noise. DNA supercoiling has been previously shown to regulate the *c-myc* proto-oncogene. We have developed genome-wide maps of transcription generated dynamic DNA supercoiling *in vivo*. We observe, experimentally, that most of the torsional stress is located within about $\pm 2000\, bp$ of transcription start site, and is differentially regulated by topoisomerases I and II.

At the RNA level, we have made an attempt to define the state of the cell using the expression levels of a sub-network of differentially expressed human kinases. Based on this definition, we have been successfully able to cluster together different molecular subtypes in lung cancer cell lines. We were able to identify and confirm previously known deregulated kinases. Many kinase genes are also identified as novel therapeutic targets. Currently we are testing these predictions, and working towards defining the complete state of a cell by getting a digital count of mRNAs

at the single cell level.

At the protein level, we studied the dynamics of protein decay to test the hypothesis that protein decay is a one step stochastic process. In several cases we have observed potentially multi-step decay processes in the ubiquitin proteasome system, however more experiments are needed before making any inferences.

# BIOLOGICAL CONTROL OF NOISE

by

Ashutosh Gupta

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Arpita Upadhyaya, Chair/Advisor
Dr. David L Levens, Co-Advisor
Professor Christopher Jarzynski
Professor Wolfgang Losert
Professor Ian White

# Preface

"So there is *DNA* inside the *nucleus* of the *cell*," Dave[1] was explaining me the project, "and we want to..."

"Wait a minute!," I pleaded, "You said three things, *DNA*: I have heard about it, but have no clue of what it is, *nucleus*: the only nucleus I know is that of the atom, and *cell*: the only cell I know is my cellphone. Lets start from there!"

This was our first discussion, when I first met Dave, my thesis advisor, late in the Spring of *2006*. It has been a long but beautiful journey.

Dave had been looking for help with analysis of microarray data for some time. He had earlier taken help from one of the senior bioinformatic professional in another reputed lab, who performed some of the routine exploratary analysis and concluded that the data was junk.

Dave found it difficult to believe that the data was junk. "When we looked at the data in the genome browser, everytime the data wiggled in interesting places," he said, "it was difficult to believe that this was junk." So he approached Dr. Wolfgang Losert at UMD seeking help with the analysis. Dr. Losert passed on this opportunity (along with others) to first and second year physics graduate students who were looking for summer positions. My journey had started taking shape.

By the end of the summer, we realized that data had significant information, and was not junk. Several new sets of experiments were performed, and eventually, as described in chapters 5 and 6, we found that the data was actually of very good quality, but needed careful attention and a novel approach.

[1] i.e. my advisor, Dr David L Levens, who prefers that everyone in lab call him Dave.

The reason why the exploratory analysis, performed previously, didn't work was that that analysis was developed for a different kind of experiments where the signal to noise ratio (SNR) is much higher. Our experiments had very low SNR, so they needed special attention and hence new special methods. Also, apart from computational approaches, a reasonable understanding of DNA physics and cell biology was needed.

If we look at it in a larger context, we see that biology is going through unprecedented changes. Last decade has seen tremendous efforts in moving biology from a qualititive science to a quantitative science.[2] On one hand, single molecule techniques (like confocal microscopy) and tools (like optical/magnetic tweezers) are putting numbers on physical properties (like rigidity/strength) of biological molecules (like proteins/nucleic acids). On the other, high-throughput techniques (like microarrays, next-generation sequencing) are enabling the amassing of massive amount of information that allows the deciphering of gene and interaction networks.

We are no longer talking about only two or three fold enrichments, but paying attention to smaller variations and corresponding statistical errors. Instead of talking about one gene, or one transcription factor, we are now probing about the gene networks (or sub-networks) of the whole organism with *100*s to *1000*s of constituents.

The field of *Systems Biology* has emerged, and is destined to revolutionize

[2]Between the announcement of the first draft of the human genome in year *2000*, to completion of this thesis work (in year *2012*), this project started right in the middle of this massive revolutionary movement (in year *2006*).

every aspect of biology we have known since the word *biology* was coined in the year *1802*.[3]

This movement from an individual to the network of individuals is not exclusive to the field of biology. A new age has begun - *the network age*, and the buzz words *cloud-mobile-social* are its clarion calls heard in our daily conversations. The first big step towards this was taken with the establishment of the *world wide web* in the year *1990*. *The web* reached a point of inflexion in the years *1998-2000* with *Google* literally becoming the common man's crystal ball for the massive network of webpages.

Massive amounts of new data are being generated everyday, data that is much larger in content than that of the Library of Congress. To analyze these massive datasets, a new field of *data science* is also taking shape at the boundaries of physics, computer science and mathematics.

Interestingly, the advent of the network age for biology had a similar timeline. Arguably the first gigantic step, marked also in year *1990*, was the announcement of the *Human Genome Project* and then its completion in year *2000*. The field has not looked back ever since. One after the other, high-throughput techniques are being developed, outpacing the theoretical developments.

When new techniques are developed, or when old techniques are used for new or customized experiments, it is desirable to develop analysis techniques to meet the computational challenge.

[3]By John-Baptiste Lamarck and, independently in the same year, by Gottfried Reinhold Treviranus and Lorenz Oken [1].

In this report, we have examples of both the cases.

The first part (chapter 2 to 6), deals with the project to develop genome-wide maps of DNA supercoiling. To generate these maps we used the old technique of microarray hybridization, but instead of ChIP-chip we used psoralen intercalation to mark superoiling of DNA (*in vivo*) and hybridized the DNA from gel purication. Psoralen has a slightly higher affinity for binding to negatively supercoiled DNA as opposed to the relaxed DNA; as a result the data is very noisy. We have developed a method to charaterize this noise, and were able to extract signal. This data was used to make inferences that were tested by an independent set of experiments. For more details see chapter 5.

In the second part (chapter 7), we have used a totally new technique, NanoStrings, to analyze the mRNA expression of human kinases in lung cancer patients. We developed a novel method to analyze the data, and were successfully able to predict the cancer types of unknown samples based on the known samples. We were also able to identify a couple of hundred new genes that were previously not known to contribute towards oncogenesis.

Both these projects demanded a lot of knowledge not only about the biology and physical systems, but also computational and programming skills. I believe that as biology becomes more quantitative, this trend will become more and more common. Just like physicists had to start learning a lot of mathematics and computational methods during last 3 centuries, the same will become a mandate for biologists in this new emerging era.

For me the journey was in the other direction, towards biology. Now when I

remind Dave about our first conversation, he recalls, "My heart just sank". But he also says that he is pleasantly surprised and happy with the progress I have made over last few years. From that first conversation in the Spring of *2006*, to the writing of this dissertation (Spring *2012*), it has been a long journey, with a steep learning curve. In retroscpect, it has been an exhilarating experience of learning and growth and I hope to continue this journey in the future.

Ashutosh Gupta

*April $4^{th}$, 2012*

# Dedication

To my mother, *Smt. Sita Devi Gupta*:
Who guided me in every tough fight...
My inspiration, my beacon of light...

To my father, *Sri Gokul Narayan Gupta*:
Who taught me how to live life...
To be humble, smile and strife...

To my darling little sister, *Jyoti (Inki)*:
My confidant, competitor and friend...
Joy of my life - until the very end...

To my dear brother, *Basant Gupta*:
Who carried responsibilities, both - his and mine...
O'brother... may you always be happy and shine...

# Acknowledgments

I owe my gratitude to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

First and foremost I'd like to thank my advisor, Dr. David Levens for giving me an invaluable opportunity to work on challenging and extremely interesting projects over the past six years. He has always made himself available for help and advice. It has been a pleasure to work with and learn from such an extraordinary individual.

I would also like to thank my UMD advisor, Professor Arpita Upadhyaya. Without her support, this thesis would have been a distant dream. Thanks are due to Professor Wolfgang Losert also, for introducing me to Dave and this wonderful opportunity.

Special thanks to all my committee members, Dave, Arpita, Wolfgang, Professor Ian White and Professor Christopher Jarzynski for agreeing to serve on my thesis committee and for sparing their invaluable time reviewing the manuscript.

My colleagues at the Levens lab have enriched my graduate life in many ways and deserve a special mention. Fedor Kouzine helped me at every step of the project and I am deeply indebted to him for being there throughout the project. The supercoiling project, and hence this thesis, would have not come to fruition without his initiative and insights.

During my initial years, Lawrence Benjamin taught me a great deal of biology. Discussions with Larry gave me lot of comfort and confidence. Really missed him during the final years when he had to take leave for health reasons.

Laura Baranello joined the lab during the final phases of the project. She helped me a great deal both in terms of helping me meet the paper deadlines, and encouraging me to finish. She is a great team player, amazing person, and a wonderful artist and scientist.

During various stages of research I have learnt a lot from other members of the group, specially Weixin Zhou, Suzzane Sanford, Zuqin Nie and Louisa Ho. I am also thankful to Hsin-Hao (H. Timothy Hsiao) and Paul Myers for their support, friendship and stimulating discussions.

The second part of the thesis (NanoString analysis for mRNA expression) was done with Dr. Avi Rosenberg, a pathologist resident. My discussions with him have been very enriching and beneficial, both scientifically and philosophically. I also thank Dr. Mark Raffeld for his support in this project.

I would also like to acknowledge my collaborators from the other projects that were a part of the research project but are not presented in this dissertation in much details: Peter Kim and Richa Rikhy from Dr. Jennifer Lippincott-Schwartz's lab (protein decay experiment); Veena Kapoor and William Telford from NIH sorting facility (protein decay experiment); my friend Monika Deshpande from Dr. Patricia Becerra's lab (cell competition experiment); my friends Erica Stein, Gema Martin Manso, David Soto-Pantoja and Tom Miller from Dr. Dave Roberts' lab (for cell competition experiment), Yvona Ward and Ross Lake from Dr. Kathy Kelly's lab (for cell competition experiment).

My thesis work was part of the first batch of collaborative projects between NCI and UMD. In absence of a precedence, it took great efforts to formalize the

details and paperwork from members of NIH and UMD staff. I would also like to acknowledge help and support from some of the staff members: Jane Hessing at UMD and Tonya Staley, Dena Flipping and Susan Hostler at NIH.

I would like to thank the Intramural Research Program of the US National Institute of Health, National Cancer Institute, Center for Cancer Research for supporting all the projects. Also, I would like to thank University of Maryland, for external fellowships to support all the course work.

My housemates at my place of residence have been a crucial factor in my life outside the laboratory. My batch mates from IIT Kharagpur Kaushik Mitra and Sandeep Mitra were always my strength. Satej Chaudhary was my roommate, friend and mentor, and continues to guide me through today long after his graduation. After my advisors, I have probably learnt the most from Satej and I am deeply indebted to him. Deepa Anagondahalli has been my roommate for the longest duration in UMD and has become an integral part of my memories here. Madhura Joglekar became a very close friend during her short stay at our place. My deepest gratitude for her friendship during some of the toughest parts of my PhD, and for her constant support towards the end of my stay here. She also helped me prepare this manuscript. I'd also like to express my gratitude to Abhinav Nigam, Ashish Mishra, Avinash Sahu, Baladitya Suri and Puneet Sharma for their friendship and support.

The on-campus Indian student group, DESI (Develop, Empower and Synergize India) was an integral of my social life and support from the first day of my arrival in USA. I would like to thank each of the members, mentors and colleagues:

Arun Shankar Mampazhy, Ajay Joshi, Prof Aravind Srinivasan, Ashwin Aravindakshan, Ashwin Kumar Kayyoor, Prof Inderjit Chopra, Kaushik Mitra, Lavanya and Om Deshmukh, Monika Deshpande, Narayanan Ramanathan, Neeraja Dashaputre, Pradeep Pandurangi, Prashant Bhoot, Priyadharshini Gowtham, Sandeep Somani, Saurabh Jain, Sharmishtha and Vinay Kelkar, Srinivasan Parthasarathy, Umang Agarwal, Utsav Chakrabarti, Vibhash Chandra Jha, Vidyaramanan Ganesan, Vijayakala Vydeeswaran and Vinod Sangwan. Each one of you have taught me so much; my experience at UMD would have been so incomplete without you.

I also received a lot of support from the local community. In particular I would like to remember and thank the following from the local community for their blessings and invaluable support: Neelam and Vinod Patel, Sonal and Satej Chaudhary, Jaisri and Ubrani Jayaram, Darsana and J. R. Josyula, Dr. Siva Subramanian, Prakash Hosadurga, Sachhidanand Babu, Keyur Patel, Bhavesh Patel, Sant Gupta, Professor Radhey Shyam Dwivedi, Professor Balaji Hebbar, Suresh Shenoy and San Sengupta.

A few weeks before my scheduled defense, my car was totaled in an accident. At that dire moment Rama and Chelakara Shankar came forward to offer me their car so that I can continue to focus on my studies. My heartiest thanks to you for this generous help and inspiring gesture.

The last few years in Maryland have been very enriching for me not only scientifically and socially, but also spiritually. My humble salutations to all my spiritual masters whose teachings have benefitted me deeply, and continue to give me strength: His Holiness Sri Sri Ravishankar, His Holiness A. C. Bhaktivedanta Swami

Prabhupada, His Holiness Swami Chinmayananda, His Holiness Pramukh Swami Maharaj, Swami Dheerananda, Prabhu Sankirtan Yagya Das (Steve), Viveknidhi Swami and Ghanshyam Sewa Swami.

My deepest gratitudes and thanks to my parents for always standing by me, guiding me, and pulling me through against impossible odds time and again. Words cannot express my gratitude for them.

I would also like to thank my entire family and, in particular, my brother Basant Gupta, who has carried out all my responsibilities in my absence back home. It is difficult to imagine me leaving home without his encouragement and presence.

Lastly, I thank the Divine, for blessing me with so much joy and to put me amidst so many wonderful people who continue to enrich my life in innumerable ways. It is impossible to recognize everyone here, and I apologize to those I've forgotten to include.

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| DNA | Deoxyribonucleic acid |
| RNA | Ribonucleic acid |
| ChIP-chip | A technique combining Chromatin Immuno-Precipitation (ChIP) with microarray technology (chip) |
| $nXL$ | non-crosslinking or non-crosslinked |
| $XL$ | crosslinking or crosslinked |
| TSS | Transcription Start Site |
| TES | Transcription End Site |
| NCI | National Cancer Institute |
| NIH | National Institutes of Health |
| UMD | University of Maryland, College Park |

# Chapter 1

# Introduction

## 1.1 Noise in biological Systems

Biological systems are remarkably precise in a lot of different ways. It is very evident from the fact that all biological entities, uni-cellular or multi-cellular, reproduce (themselves) by the very definition. In any cohort of a given type, there are remarkable large scale structural and functional commonalities that are very obvious. For example, if we consider a synchronous population of any given cell type, we will notice that they are not only structurally similar, but can respond to remarkably different stimuli in more or less similar ways. All human beings (and animals) have a certain proportion to their body parts, and are remarkably similar in the functioning of their bodies which are gigantic biological machines with trillions of living cells.[1]

Not only does the body[2] have the capacity to reproduce, it also has the capacity to defend itself from external factors. The capability to fight diseases, in particular the immune system, is an integral part of the evolution and natural selection in all plants and animals [3]. For most species there are multiple layers of defense, which

[1] Each of these cells running on a different time in their cell cycle, but somehow connected to the central circadian rhythm of the body [2].

[2] Body of any living entity, from a living human being to single bacterial or plant cell.

is adaptive and has mechanisms to remember previous attacks and successively improve the response (adaptive immunological memory) [4].

Healing or repair of damages is another peculiar hallmark of precision in biological systems. From repair of damaged DNA [5] in any single cell to healing of wounds in any part of the body, there is a constant and seemingly autonomous system of surveillance and servicing.[3]

One remarkable example of repair is liver regeneration which has been known for ages. E.g. the ancient Greeks seem to have recognized liver regeneration in the myth of Prometheus. When Prometheus steals the secret of fire from the gods of Olympus, he was punished to having a portion of his liver eaten daily by an eagle. His liver would be regenerated overnight, thus the eagle will have food eternally and Prometheus will have eternal suffering [6, 7]. It has been shown by partial hepatectomy in rats, in which specific liver lobes (amounting to about two-thirds of the liver) of a rat is removed, with the lobes left behind being intact. Within five to seven days, the residual lobes grow to make up for the mass of the removed lobes, while the removed parts of lobes do not grow [8]. When the liver from large dogs is transplanted into small dogs, the liver size gradually decreases until the size of the organ becomes proportional to the new body size [9].

From reproduction to defense and maintenance, each organism constantly monitors its internal and external environment at several different levels. Several crucial constituent factors are required to be maintained at close tolerances. A deviation, or a push, away from the equilibrium could prove fatal to an individual cell or

[3]From a nano-meter scale to meter scale, i.e. about 10 orders of spatial magnitude.

the whole organism. These deviations also have a shared history with our evolution in the form of diseases like *cancer* [10].

So how is that balance maintained?[4] With the cytoplasm of every single cell being a soup of molecules interacting stochastically, how do we get such remarkable precision at cellular and organism level? How can such seemingly improbable events occur so ubiquitously that most of us fail to even notice?

There are several known and (mostly) unknown mechanisms that govern the stochastic interactions of bio-molecules. Layered and highly integrated circuitry of various constituents (DNA, RNA, proteins, water, lipids, metals, acids, bases etc.) regulate various aspect of cellular and organismal life cycles. In this study, we present some of our efforts to understand the origin and control of this biological noise at four different levels:

1. DNA level : Developed genome-wide maps of DNA supercoiling

2. RNA level : Developed algorithm to predict cancer cell and tumor type, and predicted novel therapeutic targets from analysis of mRNA expression using NanoString technology (experiments and analysis)

3. Protein level : Studied dynamics of protein decay to test the hypothesis of one step random decay of proteins

[4]Well, for the most part of most living organisms, and for all organisms before dying, i.e. before going off balance. One may argue that death is also a part of life, but as an optimistic scientist, I would like to remind them of the famous Woody Allen quote, "I don't want to achieve immortality through my work... I want to achieve it by not dying."[11]

4. Cellular level : Explored the possibility of cell competition in mammalian cells

The entire study of this dissertation has its origins linked to a proto-oncogene called *c-myc*, which is one of the main focuses of our lab. Myc[5] is believed to regulate about 10% of entire mammalian genome.[6] It controls all major decisions of cells, including cell division and cell death, and it is known to be deregulated in most types of cancers. A noisy Myc transcription can have disastrous effects, and cells must control Myc levels very tightly. (A comprehensive review of Myc's functionality can be found here [13].)

The following sections provide a brief overview of our approach, which is expanded in subsequent chapters.[7]

## 1.2 Case for DNA supercoiling

DNA for long has been considered a passive storage house of genetic information that is acted upon by other bio-molecules. As soon as the helical structure of DNA had been drawn, understanding how the DNA strands, which intertwine around each other, are separated during DNA replication or transcription was an open and fundamental question. This task appeared to be even more challenging after the discovery of circular DNA molecules [15]. The solution used by the cell to overcome the topological problem was revealed with the discovery of DNA topoiso-

[5]Conventionally, '*c-myc*' refers to the gene, while 'Myc' referes to the protein.

[6]Our lab has recently discovered that Myc is a universal amplifier of gene expression [12].

[7]For a beautiful overview of the evolution of research in the realm of DNA, RNA and proteins during last century, see the magnum-opus [14].

merases that catalyze changes in the linkage of DNA strands and modulate DNA topology [16]. It is now certain that all DNA transactions involve alterations in the structure of DNA. The structural changes that distort the double helix through overtwisting/undertwisting and associated loop-like plectoneme structures are referred to as DNA supercoiling or DNA torsional stress[8] [17]. *In vitro* and *in silico* studies have shown that DNA supercoiling modulates the probability of DNA melting,[9] affects DNA-protein interactions, and increases the local concentration of distal DNA sites [18]. Consequently, the activities that induce DNA supercoiling may be exploited in regulatory pathways.

To understand the effect of supercoiling it is important to first place the corresponding energies and forces in the context of the other well known quantities. By means of several experimental measurements, we know that biologically relevant forces vary over a large range [19]. Thermal fluctuations and entropic forces are in pN range (energy is about $1kT = 4pN.nm$).[10] Some powerful molecular motors produce forces in the range of tens of pN (corresponding energies are e.g. $>5\,pN.nm$ (i.e. about 1–2 $kT$) for *Escherichia Coli*[11] RNA polymerase, $\sim$2–3 $kT$ for phage T7 DNA polymerase) [19]. Noncovalent interactions are in the hundreds of pN (for various hydrogen bonds, energy is about 3–12 $kT$) while covalent bonds have forces of thousands of pN (and energies in $100s$ of $kT$s) [19, 20, 21]. The energy currency of biology is ATP hydrolysis, which corresponds to about 7–10 $kcal/mol$ (about

[8] Chapter 2 reviews the basic concepts involving DNA supercoiling.

[9] Separation of the duplex DNA strands is referred to as melting.

[10] At room temperature 25 °C ($298K$), $1kT$ is equivalent to 2.479 $kJ/mol$ or 0.593 $kcal/mol$.

[11] Commonly referred to as *E. Coli*, a bacteria.

11–16 $kT$).

The energy content of supercoiled DNA can vary tremendously. For plasmid DNA[12], it increases parabolically with increasing positive or negative supercoiling (see equation 2.4). As DNA becomes more and more supercoiled, it takes increasing amounts of energy to introduce more supercoiling. To better understand this, let us consider a specific example of a 3000 $bp$ long plasmid with a superhelical density $\sigma = -.05$,[13] or a $\Delta Lk = -15$.[14] Using equation 2.4 we can find out the corresponding free energy $\Delta G_{\Delta Lk=-15} = 52.5\, kcal/mol$. If the linking number was to increase to $\Delta Lk = -16$ (or decrease to $\Delta Lk = -14$), the corresponding free energy would be $\Delta G_{\Delta Lk=-16} = 59.7\, kcal/mol$ (or $\Delta G_{\Delta Lk=-14} = 45.7\, kcal/mol$). This corresponds to a change of about $7\, kcal/mol$ ($\sim 12\, kT$), which is equivalent to hydrolysis of an ATP molecule. This shows that DNA supercoiling can store energy, and small changes (such as $\Delta Lk/N = \pm 1/3000$ above) in supercoiling can serve as the necessary energy source / sink when coupled with other reactions.[15]

Note that this energy $7\, kcal/mol$ (or $12\, kT$) is distributed over the entire plasmid at about $2\, cal/bp/mole$ (or $4 \times 10^{-3}\, kT/bp$), which seems very small. During transcription, however, if the translocation proceeds without pauses, then the RNA polymerase could generate up to 10 supercoils per second and up to 3000 supercoils for a typical 30 kbp gene [19, 23]. For actively transcribing genes, tandem initiations

[12] Plasmids are closed circular DNA molecules. Much of the earlier research was done in artificial plasmids or natural plasmids (such as E. Coli chromosome).

[13] This is typical superhelical density observed in the E. Coli bacterial genome [22].

[14] The quantities like superhelical density and linking number are defined in chapter 2 in detail.

[15] Fig. 2.9 plots the free energy, $\Delta G_{\Delta Lk}$, as a function of $\Delta Lk$ for this example.

can create large enough torsional stresses to melt the DNA. The variation in AT and GC basepairs' H-bond pairing energies (4–9 $kcal/mol$) and base stacking energies (4–15 $kcal/mol$) can facilitate the melting of DNA in a sequence dependent manner. It takes about $-9\,pN.nm$ torque to melt the DNA [24, 25, 26].

Specific melting sequence(s) could be strategically located in regions upstream (or downstream) of the transcription start sites (TSS) which can have a transcription dependent conformational change. These changes could elicit further action from other activating or repressing factors, providing a very powerful dynamic control mechanism for regulating transcriptional noise in a transcription dependent manner.

As stated before, Myc is a crucial regulator of a large number of cellular processes and a noisy Myc transcription can have disastrous effects, and cells must control Myc levels very tightly. It has been reported that a variation in *c-myc* levels could induce cell competition in drosophila melanogaster, where cells with slightly higher copy number of Myc could cause apoptosis in nearby cells with lower copy numbers [27]. In a series of papers [28, 29, 30, 31] our lab has shown that transcription generated dynamic supercoiling plays a crucial role in the regulation of the Myc proto-oncogene. (See section 3.6 for more details.)

We anticipated that the same could be true for other proto-oncogenes and regulators. As a first step in understanding the behavior of large number of genes, we decided to generate genome-wide maps of DNA supercoiling *in vivo* on ENCODE regions [32]. These regions are selected with an aim of collecting all functional elements in the human genome, and constitute about 1% of the entire genome with representation from a wide variety of gene networks and pathways. We were able

to make inferences about the large scale distribution of DNA supercoiling as well as its regulation by means of topoisomerases I and II. We were also able to show that most of the transcription generated supercoiling is confined to within about $\pm 2000\, bp$ of the TSS. We were able to separate the effects of transcription generated supercoiling from the inherent supercoiling of chromatin.

However, these maps had only about 900 genes. Many of these genes were overlapping or were poorly covered on microarray, bringing down the number of total analyzed genes to about 450. We are now planning to repeat this experiment on high density promotor arrays containing all the known genes on human genome. This will give a complete picture of the role of transcription generated dynamic supercoiling in all the genes. We would like to pay particular attention to the key regulators and oncogenes.

It is possible to generate a basepair resolution genome-wide map of DNA supercoiling on the entire genome using the 2nd generation sequencing. However, currently these high throughput experiments are prohibitively expensive for the task. With advances in technology [33, 34, 35] these costs are expected to come down and we hope to be able to develop very high resolution maps of DNA supercoiling, which will enable us to more closely examine this crucial regulator of biological noise.

DNA supercoiling research constitutes a majority of this dissertation.

## 1.3 Case for mRNA expression

RNAs are the link between the transcription and translation processes. Noise at the transcription level is propagated to translation level by means of RNAs. Transcription level noise can be introduced in several steps during the process of transcription, e.g. chromatin opening, initiation, pausing / stalling / promotor escape, elongation and termination. Apart from these, splicing, pre-processing and stability of mRNA are other critical factors for translational noise.

The second part of this dissertation focuses on this aspect of biological noise.

To study these effects, once again we started with the focus on Myc. The importance of holding Myc to close tolerances was stated in the previous section. Our aim was to get a digital count of the number of *c-myc* mRNAs at single cell level. We developed a transcript counting scheme to get a digital count of the number of Myc mRNAs in a single cell. Using some statistical methods, and advent of second generation sequencing techniques, we were able to expand our transcript counting scheme for all mRNAs at single cell level.

This is a powerful method for understanding transcriptional / translational noise as well as for probing the origins of the noise by inferring functional relations among various genes. The complete transcritome can also serve as the definition of the state of a single cell, or a population of cells. These states can then be compared between diseased and healthy cells / populations.

The project has many challenges, e.g. isolating single cells, extracting mRNA from single cells, maintaining enzyme activities for various reaction buffers etc. Dur-

ing our pilot studies we were able to extract Myc from single cell levels. Preparations are now on for completing these experiments.

For this dissertation, we will present a small scale variant of the transcript counting experiment described above. Instead of counting all the mRNA transcripts from a single cell, we can get an estimate of mRNA copies of a set of pre-selected transcripts using the NanoString nCounter$^{TM}$ assay system [36].

The highly sensitive NanoString nCounter system is useful for a variety of applications, such as digital counting of miRNA and mRNA transcripts across a dynamic range and measuring copy number variation of DNA. However, the high sensitivity may cause large distortions in data due to experimental variables such as small variation in sample preparations and loading, as well as non-specific binding of some probes. A novel normalization and error correction approach was developed utilizing the built in "stable" house-keeping genes along with the positive and negative controls. In this preliminary report, analysis of NanoString data is presented using the novel protocol to normalize data for a set of 22 lung cancer cell lines (controls and treatments) on Human Kinase codeset.[16] The data is analyzed in various ways to get significant and useful insights about the clustering of the various molecular subtypes of lung cancer and the functional information about various targeted drugs and kinase genes that are affected.

[16]Kinases are a type of transfer enzymes that catalyze the transfer of phosphate groups from high-energy donors, such as ATP, to specific acceptor molecules (a process known as phosphorylation) [37]. This codeset had markers for 519 (out of a total of $> 2000$) human kinases.

## 1.4 Case for protein decay

Just like the transcriptome, the proteome[17] can define the state of the given cell or organism. Noise in the proteome is originated during transcriptional and translational processes, and this noise is fed back to the transcriptional / translational processes. Once the protein is made, variation in its activity, mode of movement (e.g. diffusive, distributive, processive), stability, and decay are key contributors that cause variation in its functional output in the pathways downstream.

One of the central contributing factors to this process is protein stability and decay. It has been believed for a long time that protein decay is a one step stochastic process. We believe that a one step stochastic protein decay would be very noisy and could have many undesirable effects. This would be particularly true for the case of proteins like Myc that have small number of copies[18], $< 500$ in resting fibroblast cells, and a small half-life, $t_{\frac{1}{2}} \sim 30\,min$. As discussed in the previous section, it has been reported that cells with slightly higher copy number of Myc could cause apoptosis in nearby cells with lower copy numbers [27]. This means if a cell has too many more copies than its neighbors, it will start killing them. Or, if it has less copies than the neighbors, then it might get killed.

So random disappearance of crucial 'life support field workers', like Myc, could be fatal to the cell or its neighbors, and to the organism for sure. Cells must have processes to prevent this from happening. One such possibility is a multi-step decay process, with some sort of signaling or feedback indicating nearing 'resignation' of

[17] Proteome is the set of all the proteins expressed by the genome of a cell or organism.

[18] Just like stability is important in case of mRNA transcripts in previous section.

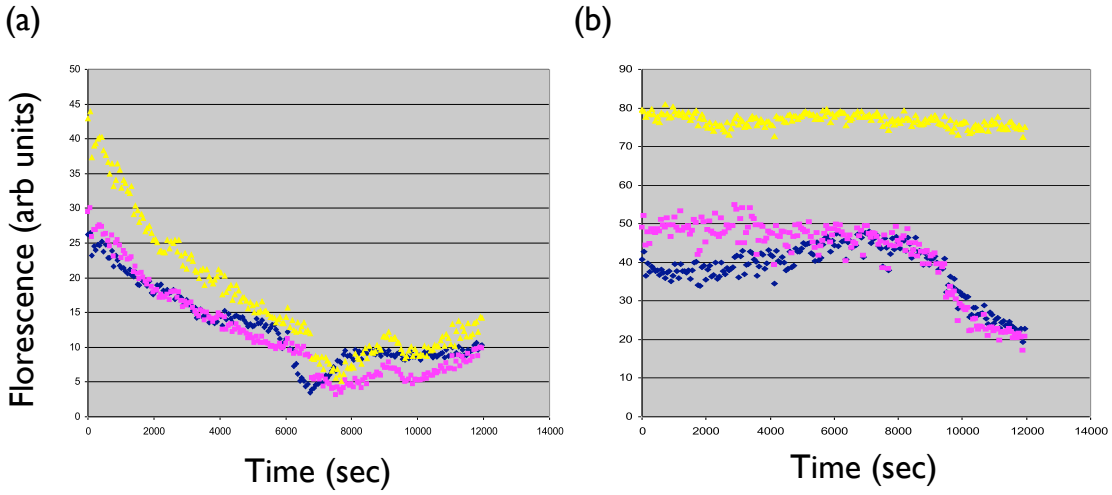


Figure 1.1: Studying protein decay: Experiments were done in triplicates (see text for more details) (a) One step stochastic decay (b) Multi-step stochastic decay

the field worker and need for a new 'recruit'.

To test the dogma of one step stochastic protein decay, we performed experiments at single cell levels, on Myc which decays using ubiquitin proteasome system (UPS) [38]. It is a pulse-chase experiment[19] on florescently tagged Myc fusion proteins, transfected in a mammalian cell line (HeLa cells, a cancer cell line). If the decay is indeed a one step process, we should see an exponential decay in the florescence. However, if the decay is not a one step process, we should observe a plateau before the decay starts. The number of kinetic rate limiting steps in the decay process can be inferred from the shape of the decay curve.

During our pilot experiments, the results showed high variability. In some cases we observed a one step decay (Fig. 1.1a), while in several cases we do observe a multistep decay - as per our prediction (Fig. 1.1b). However, a large number of the experiments showed cell death during the study. The data is not in conclusive

[19] i.e. to study the decay of a protein following the pulse of production.

or presentable format yet and more experiments are needed before making any significant inferences.

We believe that this variability is a result of extended exposure of nucleus (as Myc is a nuclear protein) to UV radiation causing DNA damage, and potentially activating apoptosis pathways. We are now planning to do more experiments using cytoplasmic proteins, which would not need exposure of the nucleus to UV radiation.

## 1.5 Case for cell competition

Lastly, at the cellular level, we studied the cellular implications of noise in DNA, RNA and Protein levels in a phenomenon known as cell competition. Cell – cell interaction is important for all stages of growth, maintenance and disease in any organism [39, 40]. Resource allocation, distribution, and inter-cellular communication are crucial for survival of all cells. Competition is an integral part of the dialogue that determines the survival as well as the status of cells. Cell competition is the phenomenon where two metabolically different populations, within the same growing tissue, confront each other and the fittest survives [40].

As stated in previous sections, Myc has been implicated in cell competition in drosophila melanogaster [27, 41]. So continuing with our theme of focusing on Myc to study the biological noise, we examined the possibility of Myc mediated cell competition in mammalian cell line (mouse fibroblast 3T3).

To study the effect, we used two cell populations. One population of cells had stable transfections of tamoxifen inducible Myc-ER fusion proteins, while the other

population had a florescent reporter protein (GFP). Myc-ER fusion proteins lose Myc functionality due to Estrogen Receptor (ER). Tamoxifen is an antagonist of estrogen receptor, and in presence of tamoxifen the Myc activity is regained. This reversible process gives us a way to generate an instantaneous pulse of high Myc levels in Myc-ER cells. This cell population while mixed with the GFP population can generate conducive environment for cell competition. Upon addition of tamoxifen, the Myc-ER cell line will have a much higher level of activity as compared to the GFP cell line. If cell competition is occurring, GFP cells should have a much higher death rate than the Myc-ER cells.

During our experiments, we observed that the cells with higher levels of Myc do have a higher proliferation rate (which has been reported before [42]) while the GFP cells, with lower levels of Myc activity, were growing at a much slower pace. However, beyond the differential proliferation rate, we did not observe any cell competition.

More experiments are needed to test the possibility of cell competition in mammalian cells. (Cell competition has not been reported in mammals so far.)

## 1.6 Outline of Thesis

A majority of our focus in this thesis remains DNA supercoiling. We start with a brief review of DNA as a molecule, DNA supercoiling and other key players in chapter 2. Chapter 3 discusses biological relevance of DNA supercoiling and gives an overview of the current literature. Chapter 4 discusses some insights into possible

ways of measuring DNA supercoiling using microarrays. Chapter 5 discusses the results and inferences from our experiments and analysis. Chapter 6 summarizes the analysis methods used in chapter 5.

Following this, chapter 7 summarizes the mRNA expression analysis and inferences using the NanoString technique on Human Kinase codeset.

Protein decay and cell competition projects need more work and are not included in this dissertation.

# Chapter 2

# Introduction to DNA mechanics and topology

In this chapter we will review some basic facts and concepts about DNA, DNA supercoiling and some of the key players that play an important role in generation and regulation of DNA supercoiling. The purpose is to have a basic familiarity with these concepts. For a detailed understanding it is advisable to see some of the standard texts [22, 43].

Most of the material in this chapter was prepared for and presented as an introductory lecture on "DNA, Torque and Cancer" at UMD, College Park (invited talk for the biophysics course PHYS 818, May 2010).

## 2.1 Introduction

DNA (or DeoxyriboNucleic Acid) is a type of nucleic acid that contains the genetic code for the respective species' growth and functioning. The genetic information stored in DNA is copied to another type of nucleic acids - mRNAs (messenger RiboNucleic Acid), which are blue-prints for manufacturing of the molecular machines, i.e. proteins. The regions with genetic information are known as genes. Other regions have regulatory, functional or structural purposes, and it is believed that about 98% of mammalian genome doesn't code for proteins [44].

DNA is made up of two conjugate strands of nucleic acids, each with a sugar phosphate back bone that has a $5'$ to $3'$ directionality. There are four bases (A-adenine, T-thymine, G-guanine and C-cytosine), one of which hangs from each subsequent sugar of the sugar phosphate backbone. The strands are held together due to the hydrophobic nature of the bases, and the 2-3 hydrogen bonds between the conjugate bases (2 H-bonds for A-T, and 3 H-bonds for G-C).

## 2.2 DNA supercoiling and linking number

The two strands of DNA are coiled together to make a double helix or a coil. When this coil is overtwisted or undertwisted from its relaxed form, it is called a supercoil. Overtwisting or tightening is referred to as *positive supercoiling*, while undertwisting or loosening is known as *negative supercoiling*.

To understand the idea of supercoiling let us consider a very simple example. Fig. 2.1a shows two circles standing alone. As they are not connected, i.e. not linked, their linking number, $Lk = 0$. Now if one of them is broken and resealed interlocking the other circle, we will get a configuration similar to Fig. 2.1b. As it takes one breaking / resealing of one circle to unlink the two circles, we say that the circles are linked once, or their linking number $Lk = 1$.

Further, note that it doesn't make any difference which of the strands were broken here. Since there is no sense of directionality to the strands, the representation in Fig. 2.1b is good enough for describing the configuration of linking number, $Lk = 1$. The other configurations are plotted in Fig. 2.2b. Both these configurations

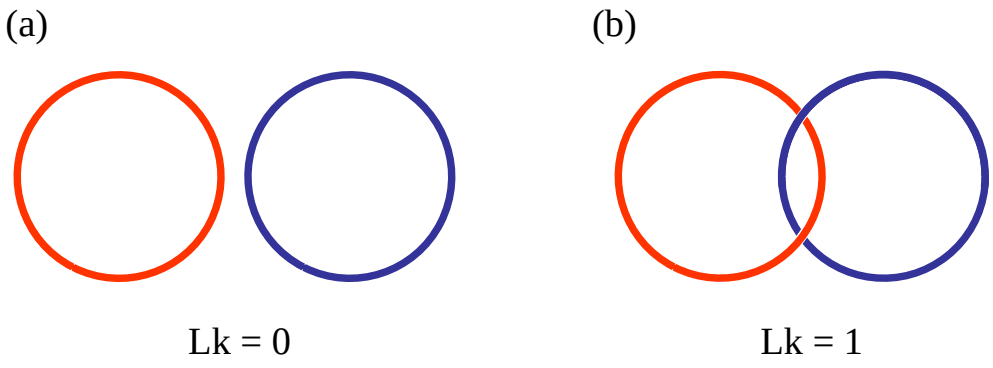


Figure 2.1: Two circles (a) When unconnected, linking number, $Lk = 0$, (b) When connected once, $Lk = 1$.

have same linking number, i.e. $Lk = 1$.

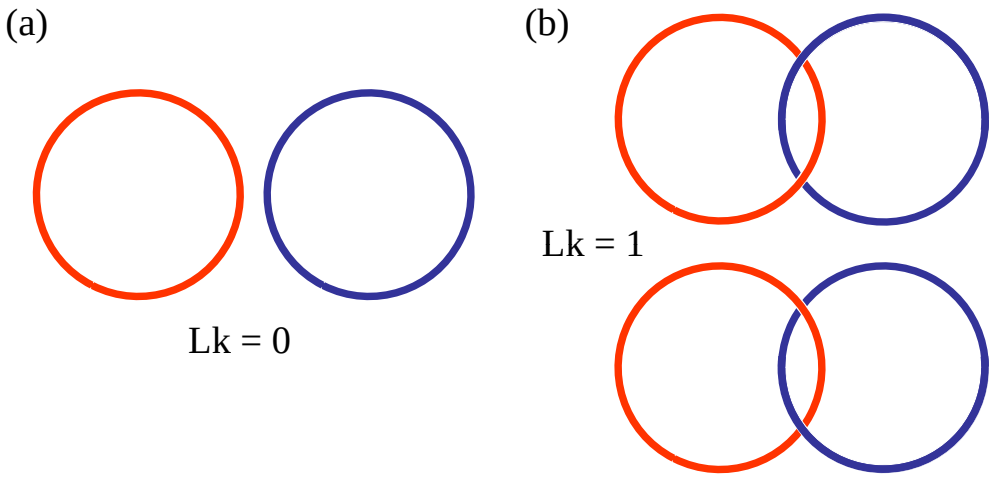


Figure 2.2: In absence of directionality different configurations of the linked circles in Fig. 2.1b are superimposable, and have the same linking number.

As stated in the previous section, DNA strands do have a sense of directionality to them. In the duplex (double helix) form, the two strands are running in opposite directions. See Fig. 2.3. Conventionally the direction from $5'$ to $3'$ is considered positive.

Due to directionality on both the circles, it is important to keep track of the

crossings. The two configurations of Fig. 2.3b are no longer superposable to each other, see Fig. 2.4.

The linking numbers of the two configurations in Fig. 2.4 have linking numbers of $+1$ and $-1$ respectively. Fig. 2.5 summarizes the rules to determine the linking number of linkage of directional strands from a $2D$ representation.

The rules can be summarized as follows. After moving over the crossing, start from the strand on top and draw an angular arrow (mentally or literally):

1. If the arrow is drawn $anti-clockwise, \Rightarrow Lk = +1/2$,

2. If the arrow is drawn $clockwise, \Rightarrow Lk = -1/2$.

After putting a number on all the cross-overs, a sum total of the these numbers gives the linking number of the overall assembly.

Note that a mere rotation of the $2D$ representation (or looking from the other side of the paper) will not change the rule. Also see Fig. 2.6.

These simple rules are suffice to understand the basic concepts of supercoiling we need for this dissertation. If the circles are linked more than once, as in Fig. 2.7, the linking numbers can be computed accordingly.

Now let us move to the specific implications of these for DNA.

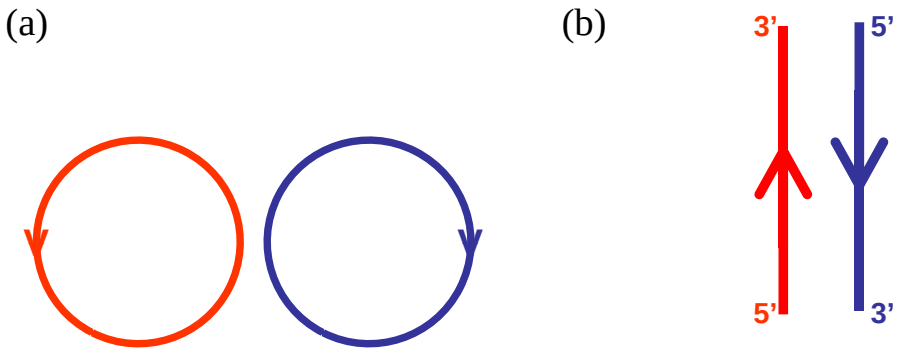


Figure 2.3: Directionality of the two strands of DNA. (a) When unconnected linking number $Lk = 0$, (b) In the duplex form two strands have opposite directions. Conventionally $5'$ to $3'$ is considered positive.

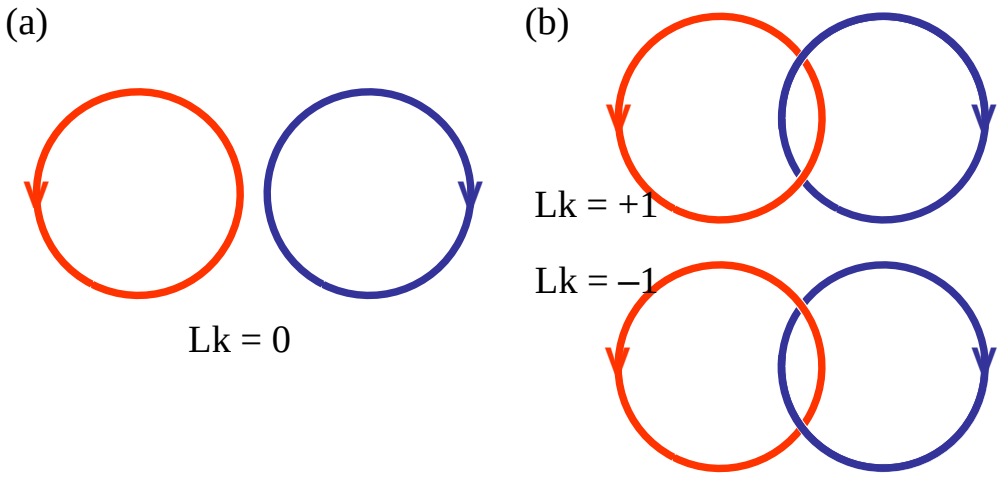


Figure 2.4: In presence of directionality the configurations of the linked circles in Fig. 2.2b have different linking numbers.

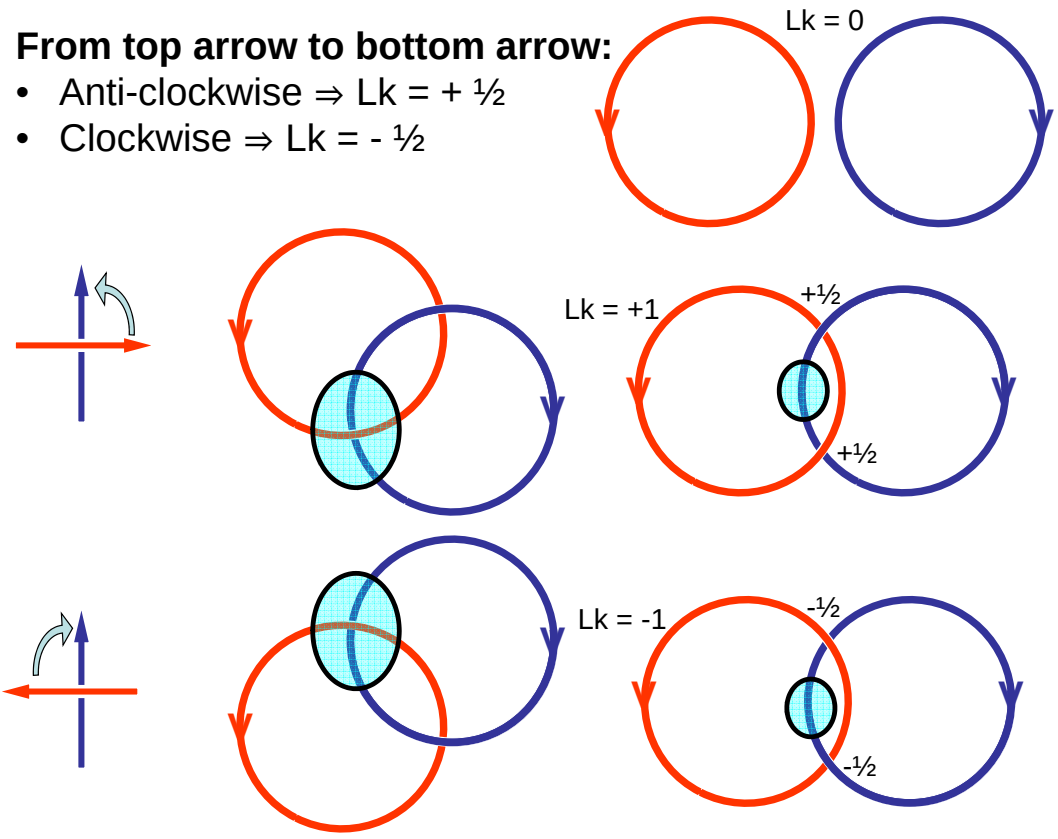


Figure 2.5: Summary of rules to determine the linking number in linked domains of directional strands from a $2D$ representation as shown in Fig. 2.4.

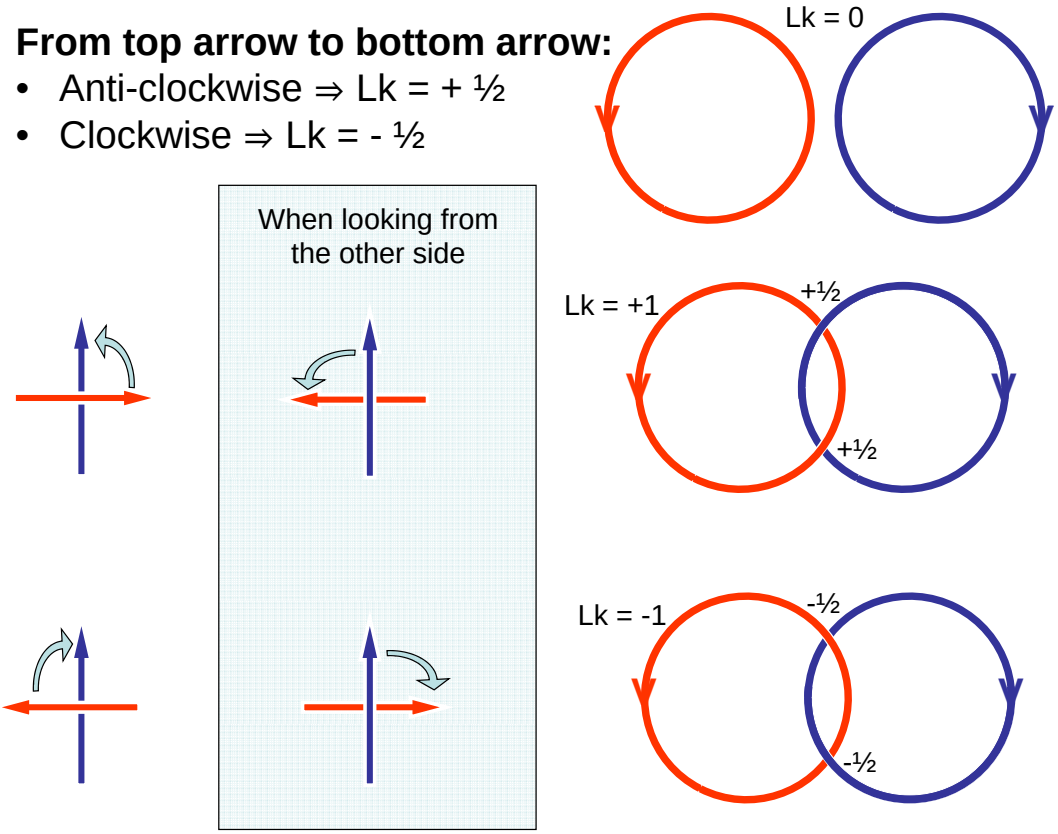


Figure 2.6: Uniqueness of rules to determine the linking number of linked domains of directional strands from a $2D$ representation.

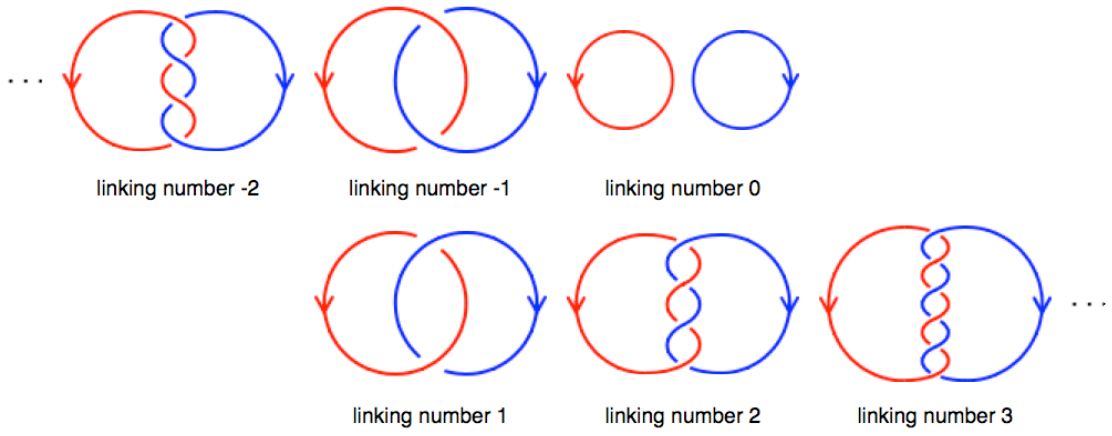


Figure 2.7: Higher linkages in directional strands as compared to Fig. 2.6. (Figure reproduced from [45] under free public license)

## 2.3 Topology of relaxed and supercoiled DNA

Fig. 2.8 shows a schematic representation of relaxed DNA at 25°C. The relaxed form of DNA is usually referred to as B-DNA.[1] Note that here the strands are running in the opposite direction in a 'structural' (or biolchemical) sense, however, topologically they are assumed to be running in the same direction. Hence the linking number is to be computed assuming same directionality.

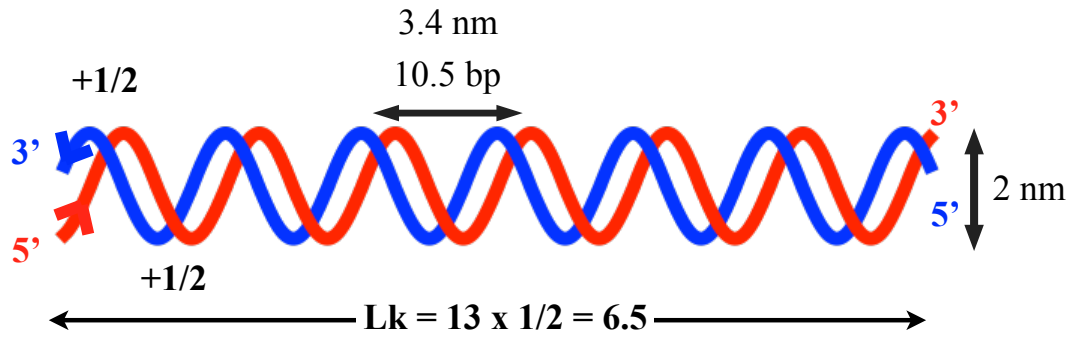


Figure 2.8: Schematic diagram of DNA with dimensions (at 25°C) along with linking number calculation.

At 25°C, one helical turn (or pitch, $h$) of DNA constitutes about $10.5\,bp$, i.e. every $10.5\,bp$, linking number increases by one. So for an $N\,bp$ long piece of relaxed DNA, the linking number is given by:

$$Lk_\circ = \frac{N}{h} \tag{2.1}$$

If the DNA is tightened or loosened, the linking number would change to a new value, $Lk$, and the corresponding change is given by:

[1] There are several other types of structural variants DNA configuration that are physiologically present [43, 46].

$$\Delta Lk = Lk - Lk_\circ \tag{2.2}$$

Our ultimate goal of this study is to be able to generate a genome-wide map of the linking number change (from relaxed DNA) due to various processes. However, we will use a more commonly used variant of $\Delta Lk$, known as superhelical density, $\sigma$, which is defined as:

$$\sigma = \frac{Lk - Lk_\circ}{Lk_\circ} = \frac{\Delta Lk}{Lk_\circ} \tag{2.3}$$

Note that tightening of DNA duplex squeezes more turns for the same length of DNA, i.e. $Lk > Lk_\circ$, therefore $\Delta Lk$ is positive, and hence tightening is known as positive supercoiling. Similarly, for loosening $Lk < Lk_\circ$, therefore $\Delta Lk$ is negative, and hence the name negative supercoiling.

## 2.4 Free energy associated with DNA supercoiling

The specific free energy, $\Delta g_{\Delta Lk}$, associated with the change in linking number, $\Delta Lk$ (or superhelicity, $\sigma$), in a plasmid of $N\,bp$, can be given by [22, 47]:

$$\begin{aligned} \Delta g_{\Delta Lk} &= \Delta G_{\Delta Lk}/N \\ &= NK(\Delta Lk/N)^2 \\ &= \frac{NK}{h^2}\sigma^2 \end{aligned} \tag{2.4}$$

$\Delta Lk$ is the change in linking number between supercoiled and relaxed state [22], $h$ is the helical repeat for relaxed DNA (or pitch i.e. $10.5\,bp/turn$, and NK

is also a constant which is confirmed experimentally for $N > 2000\,bp$ as $1200RT$ (or $700\,kcal/mol$ at $20\,^{\circ}$C), to within about $\pm 10\%$, and for $N \approx 200\,bp$ as $3900RT$ (or $2275\,kcal/mol$) [47]. Note that free energy ($\Delta G$) has a squared dependence on supercoiling density ($\sigma$), which means any deviation from relaxed state (positive or negative) would cost energy, as expected.

As discussed in chapter 1, the energy content of the supercoiled DNA increases parabolically with increasing positive or negative supercoiling. As DNA becomes more and more supercoiled, it takes increasing amounts of energy to introduce more supercoiling. To better understand this, let us consider a specific example of a $3000\,bp$ long plasmid with a superhelical density $\sigma = -.05$[2], or a $\Delta Lk = -15$. Using equation 2.4 we can find out the corresponding free energy $\Delta G_{\Delta Lk=-15} = 52.5\,kcal/mol$. If the linking number was to increase to $\Delta Lk = -16$ (or decrease to $\Delta Lk = -14$), the corresponding free energy would be $\Delta G_{\Delta Lk=-16} = 59.7\,kcal/mol$ (or $\Delta G_{\Delta Lk=-14} = 45.7\,kcal/mol$). This corresponds to a change of about $7\,kcal/mol$ ($\sim 12\,kT$), which is equivalent to hydrolysis of an ATP molecule.

This shows that DNA supercoiling can serve as a storage of energy, and small changes in supercoiling can serve as the necessary energy source / sink when coupled to other reactions.

Fig. 2.9 plots the free energy, $\Delta G_{\Delta Lk}$, as a function of $\Delta Lk$ for this case. Note that this energy ($12\,kT$) is distributed over the entire plasmid at about $2\,cal/bp/mole$ or $4 \times 10^{-3}\,kT/bp$, which seems very small. During transcription, however, if the translocation proceeds without pauses, then the RNA polymerase could generate

[2]This is typical superhelical density observed in the E. Coli bacterial genome.

up to 10 supercoils per second and up to 3000 supercoils for a typical 30 kbp gene [19, 23]. For actively transcribing genes, tandem initiations can create large enough torsional stresses to melt the DNA. The variation in AT and GC basepairs' H-bond pairing energies ($4 - 9\,kcal/mol$) and base stacking energies ($4 - 15\,kcal/mol$) can facilitate the melting of DNA in a sequence dependent manner. It takes about $-9\,pN.nm$ torque to melt the DNA [24, 25, 26].[3]

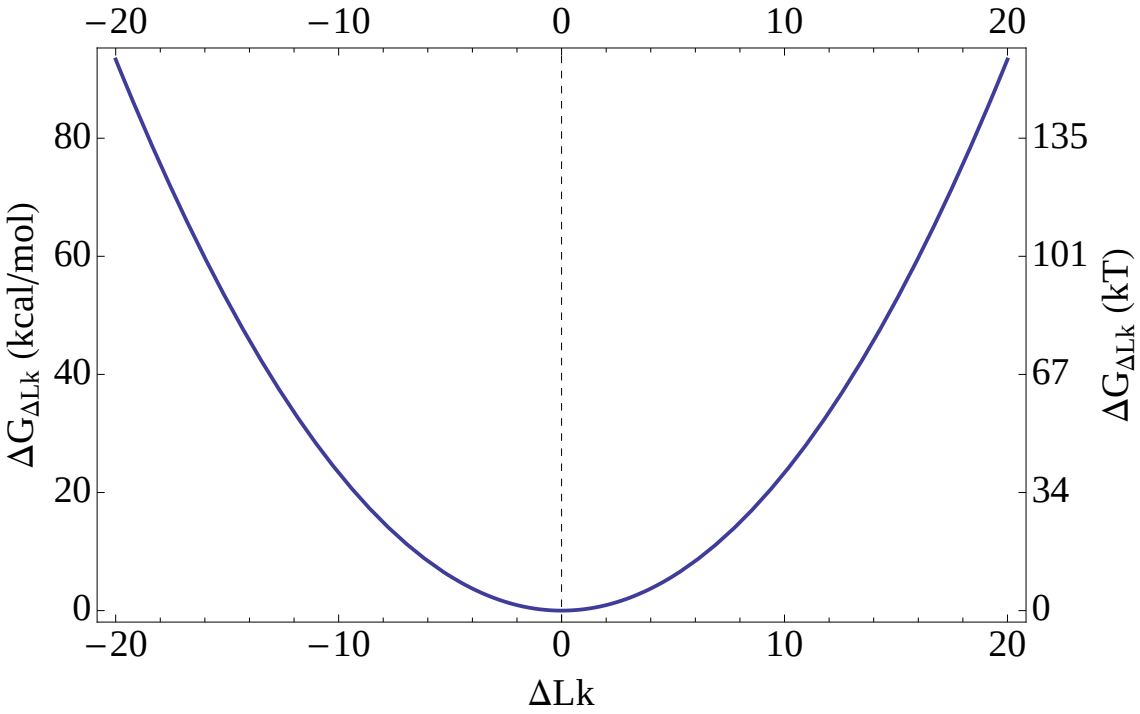


Figure 2.9: Free energy, $\Delta G_{\Delta Lk}$, as a function of $\Delta Lk$ for a $3000\,bp$ long plasmid at $20\,^{\circ}$C. See equation 2.4 for more details.

Specific melting sequence(s) could be strategically located in regions upstream (or downstream) of the transcription start sites (TSS) which can have a transcription

[3]Note that the parabolic model will break down as the material (i.e. DNA) goes through a phase transition. It will break soon for the negative supercoiling because DNA will start to melt at about $-9\,pN.nm$, while for positive supercoiling a torsion of about $20\,pN.nm$ would be needed for a transition to P-DNA form (P stands for Pauling) [24, 25, 26].

dependent conformational change. These changes could further elicit action from other activating or repressing factors, providing a very powerful dynamic control mechanism for regulating transcriptional noise in a transcription dependent manner.

## 2.5 Key players for generation and relaxation of DNA supercoiling

Although almost all DNA – protein interactions cause the DNA to deviate from its relaxed B-DNA form, there are two types of enzymes that require special attention, namely *polymerases* and *topoisomerases*.

Polymerases are enzymes that copy the DNA content for replicating DNA (hence the name DNA polymerase) or for RNA transcription (hence the name RNA polymerase). Since they thread through the DNA, without making a break, they introduce positive supercoiling downstream (i.e. in the direction of translocation) and negative supercoiling upstream. In case of fluent translocation polymerases can introduce supercoiling at about $9 - 10\,\Delta Lk/sec$ and up to 3000 supercoils for a typical $30\,kbp$ gene [19, 23].

Topoisomerases on the other hand are enzymes that relax the supercoiled DNA. They achieve this by nicking, or breaking, one strand of DNA (hence the name topoisomerase I) and passing over the other strand or by nicking both the strands of DNA (hence the name topoisomerase II), and passing of another region of double stranded DNA. For a detailed description of topoisomerases, see chapter 5 of [22].

## 2.6 Modeling DNA mechanics

As discussed in the previous sections, most DNA–protein interactions involve alterations of DNA structure from its relaxed state to a greater or lesser degree. The *in vivo* diffusion, through DNA fibers, of torsional stress generated during the process of transcription have been a matter for speculation for several decades [48]. Recent developments in single molecule techniques have confirmed that DNA does not behave like a rigid rod, in fact, not even like a plumbers snake [49]. These results suggest that even though DNA doesn't behave like a rigid rod, it might (as speculated in [43]) still need to be anchored to a structure or to be closed upon itself – so as to form a precisely bounded topological domain (otherwise in case of an open domain, the other end might rotate freely). It was proposed and experimentally shown [50] that the frictional drag acting upon DNA in a viscous aqueous medium could increase the capacity of DNA to absorb the torsional stresses and retard their diffusion.

In a closed topological domain the linking of DNA is given by:

$$Lk = Tw + Wr \tag{2.5}$$

where $Lk$ is the *linking number* (defined earlier in section 2.2), $Tw$ is *twist* – representing the coiling of the individual strands about each other (as described in section 2.2), and $Wr$ is *writhe* – representing the over all undulations, in 3 dimensions, of the central helical axis of duplex DNA.[4] In a relaxed DNA circle sitting on a plane, the writhe contribution is zero, and linking comes from twisting the two

[4]For a detailed discussion please see chapter 2 of [22].

DNA strands about each other. However, when the DNA is supercoiled, writhing provides a way to release (or redistribute) the torsional stress:

$$\Delta Lk = \Delta Tw + \Delta Wr \tag{2.6}$$

At very low levels of supercoiling the torsional stress can be accommodated as small changes in twist (reflected in slight lengthening or shortening of the helical pitch). However, with increasing levels of supercoiling the duplex begins to fold on itself with the helical axis asymmetrically shifting from the plane of relaxed DNA (anchored linear DNA or plasmid DNA), i.e. introducing writhe into the duplex.

It has been shown that RNA polymerase bound on DNA bends the duplex by 90 degree [51]. During transcription, the positive supercoiling downstream (i.e. in the direction of motion) is mainly introduced as twist. However, due to the aforementioned 90 degree bend, the negative supercoiling generated upstream of RNA polymerase is first manifested in writhe and later repartitioned partially as twist. It is easy to model torsional stress distributed in twist, however developing an analytical mathematical model for writhing is a bit challenging, although there have been several attempts to simulate the effect of supercoiling in naked plasmid DNA [18, 52, 53, 54]. *In vivo*, in the setting of chromatin where the trajectory of the chromatin fiber and the boundaries of stable and flickering topological loops are ill-defined, it is unclear what assumptions may be made to simplify the modeling.

This is a critical roadblock for our capacity to analytically model the DNA supercoiling and more work is needed to expend our understanding of the process.

# Chapter 3

# Overview: DNA supercoiling and regulation of dynamic processes

This chapter serves as literature review for the field of DNA supercoiling. The chapter was published as a review paper [55], for which I was a minor author.

## 3.1 Summary

Through dynamic changes in structure resulting from DNA-protein interactions and constraints given by the structural features of the double helix, chromatin accommodates and regulates different DNA-dependent processes. All DNA transactions (such as transcription, DNA replication and chromosomal segregation) are necessarily linked to strong changes in the topological state of the double helix known as torsional stress or supercoiling. As virtually all DNA transactions are in turn affected by the torsional state of DNA, these changes have the potential to serve as regulatory signals detected by protein partners. This two-way relationship indicates that DNA dynamics may contribute to the regulation of many events occurring during cell life. This chapter summarise the current literature and gives an overview of how DNA supercoiling plays an important role in the cellular processes, with particular emphasis on transcription. Besides giving an overview on the multiplicity of factors involved in the generation and dissipation of DNA torsional stress,

we will discuss recent studies which give new insight into the way cells use DNA dynamics to perform functions otherwise not achievable.

## 3.2 Introduction

DNA for long has been considered a passive storage house of genetic information that is acted upon by other bio-molecules. As soon as the helical structure of DNA had been drawn, understanding how the DNA strands, which intertwine around each other, are separated during DNA replication or transcription was an open and fundamental question. This task appeared to be even more challenging after the discovery of circular DNA molecules [15]. The solution used by the cell to overcome the topological problem was revealed with the discovery of DNA topoisomerases that catalyze changes in the linkage of DNA strands and modulate DNA topology [16]. It is now certain that all DNA transactions involve alterations in the structure of DNA. The structural changes that distort the double helix through overtwisting/undertwisting and associated loop-like plectoneme structures are referred to as DNA supercoiling or DNA torsional stress (Fig. 3.1a) [17]. *In vitro* and *in silico* studies have shown that DNA supercoiling modulates the probability of DNA melting, affects DNA-protein interactions, and increases the local concentration of distal DNA sites [18]. Consequently, the activities that induce DNA supercoiling may be exploited in regulatory pathways.

In bacteria, the genomic DNA is maintained in an undertwisted state which facilitates localized melting of the double helix at origins of replication or transcription

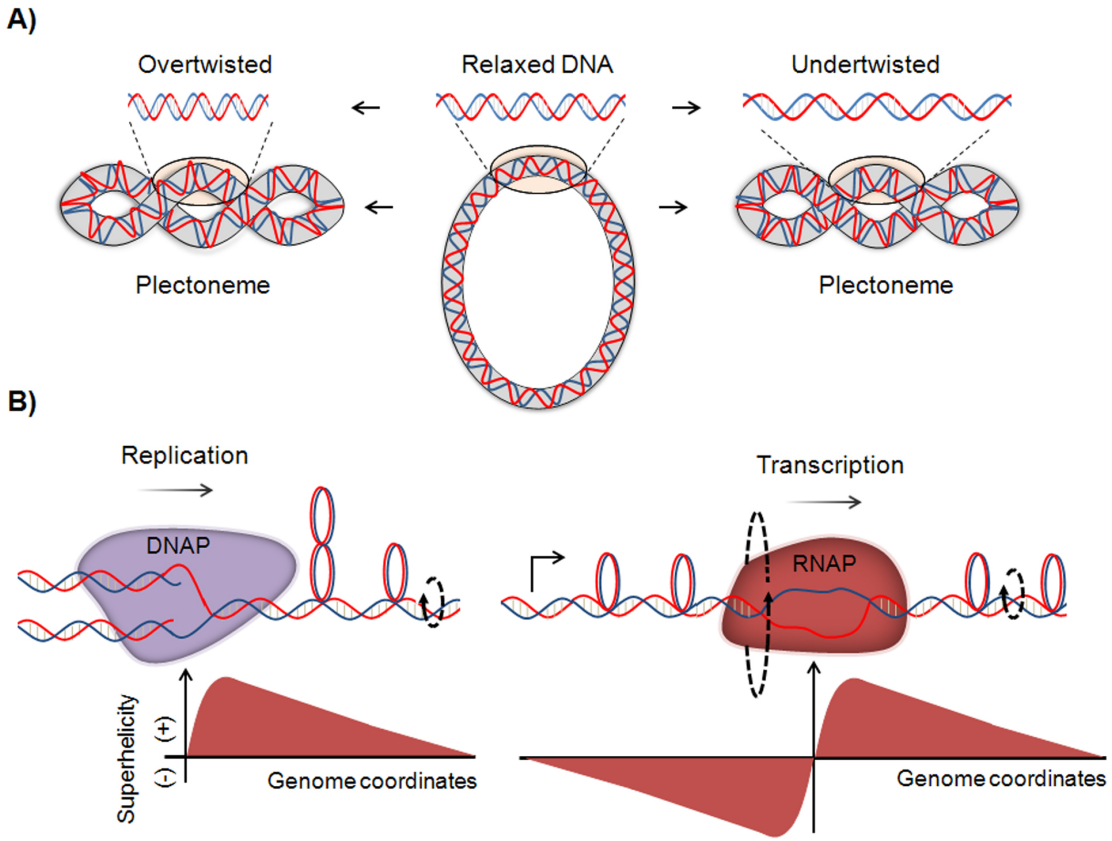


Figure 3.1: Basics of DNA topology and its relevance to DNA transaction: The DNA topology is described quantitatively by the twist of double helix and by the number of times the helix crosses over on itself (plectoneme). Plectonemic structures are typically formed by bacterial plasmids. B) A graphical illustration showing the generation of supercoiling during transcription and replication. If polymerases are moving without rotation, then due to its helical structure, the DNA must be screwed through the protein complexes. In this case, the templates rotate around its axis as indicated by curved arrows.

initiation sites, contributes to the formation of the nucleoid structure and promotes recombination events [56, 57, 58]. The concerted activities of topoisomerases and gyrases (DNA supercoiling enzymes) are determinant for maintaining the supercoiling homeostasis necessary to optimize these key genetic processes [59]. Eukaryotic organisms lack enzymes such as DNA gyrase that directly introduce supercoils into DNA, but statically their genome is supercoiled to a similar degree of bacterial genome [60]. Each nucleosome of the chromatin is wrapped by DNA 1.8 times and constrains approximately one negative supercoil which cannot diffuse to remote areas until released by nucleosome removal [61, 62]. Thus, as a consequence of the chromatin organization, the net of DNA supercoils is fixed in the eukaryotic genome and is known as constrained supercoils. The unconstrained supercoils must be accommodated within the linker DNA (regions separating the nucleosomes) which in average represents only 20% of the genomic DNA in higher eukaryotes and decreases up to 6% in the yeast [63, 64]. Dynamic interplay between broadly distributed constrained supercoils and the local unconstrained supercoils in the eukaryotic genome complicates the assessment of the DNA torsional state in the cells [65, 66, 67]. Only recently the experimental approaches have advanced to the point where it is feasible to interrogate the role of DNA topology in gene regulation.

## 3.3 Origin of DNA supercoiling

Cellular processes dynamically change DNA topology. According to the supercoiled domain model the activities that force DNA to revolve around its axis

generates a local domain of DNA supercoiling (Fig. 3.1b). This hypothesis applies with minor modification to the movement of transcription and replication complexes as well as for some helicase and restrictase activities [48, 68, 69, 70]. Currently, the best investigated example is transcription-generated supercoiling. Due to the overwhelming molecular mass of the RNA polymerase and given the arguments in favor of immobilization of RNA polymerase in transcription factories, the DNA template is forced to rotate around its axis as the double helix threaded through the transcriptional machinery [71, 72, 73, 74]. The upstream DNA becomes untwisted, while the downstream DNA becomes overtwisted which is referred to as negatively and positively supercoiled, respectively. If the translocation proceeds without pauses then the RNA polymerase could generate up to 10 supercoils per second and up to 3000 supercoils for a typical 30 kbp gene [19, 23]. This enormous torsional stress might be inhibitory for efficient transcription [48, 75, 76]. Consequently, it is relieved by DNA topoisomerases which transiently break and rejoin the backbone of DNA [69].

Another source of DNA supercoiling is provided by the reorganization of eukaryotic chromatin: the disassembly or assembly of nucleosomes releases or absorbs DNA superhelicity. Special protein complexes called chromatin remodelers are able to remove or slide nucleosomes in an ATP-dependent fashion [77, 78]. Notably, *in vitro* experiments have shown that these chromatin remodeling activities directly generate torsional stress of DNA in the presence of nucleosomes [79]. While the remodeling of the chromatin structure is a broad phenomenon that could involve sometimes entire loci, it is very difficult to assess and measure *in vivo* the extent of generated unconstrained supercoiling due to the transient nature of this process

which could be unsynchronized in a population of cells [80, 81]. Consequently, direct evidence is still needed.

In addition to DNA-tracking activities and chromosome remodelers, the existence of nuclear actins and myosin in principle may allow mechanical forces to be applied directly to chromatin fibers [82, 83]. Single DNA molecule experiments *in vitro* have demonstrated a dynamic coupling between twisting-untwisting of the double helix and stretching forces, a possibility which remains largely unexplored *in vivo* [84].

## 3.4 Tuning of transcription-generated DNA supercoling

The level of supercoiling depends on two opposite processes: how fast torsional stress is introduced into the DNA, and how fast it is relaxed or diffused into remote regions of the genome. The supercoil generation in the DNA flanking RNA polymerase complexes depends on the rate of transcriptional elongation which may be relatively invariant in the absence of specific RNA polymerase pausing or stalling and on the rate of transcriptional initiation [48, 85, 86]. Thus low level transcription produces a pulse of torsional stress followed by DNA relaxation, while high level transcription, due to repetitive initiation, may establish stable dynamic supercoiling upstream of transcription start sites [28, 31]. In the transcribed unit of highly active genes the DNA regions between RNA polymerases transcribing in tandem contain supercoils of opposite polarity that could annihilate each other. Other important parameters include the distribution of promoters which, in divergent orientation,

could reinforce DNA supercoiling upstream transcription start sites by untwisting the double helix as well as by inducing directly plectonemes [87], and the presence or absence of barriers to diffusion of torsional stress [88]. The dynamics of supercoil diffusion should depend on the behavior of chromatin fibers: in principle, the position of individual nucleosomes, the interactions between them, the linker binding proteins and the nucleosome modifications will govern supercoil propagation. We still do not know much about these important properties of chromatin, but single nucleosome array experiments *in vitro* reveal high torsional flexibility of chromatin compared to naked DNA [66, 89]. Successively, it has been found that chromatin fiber behaves qualitatively similar to the nucleosome arrays, probably due to the conformational flexibility of nucleosomes [90]. If the same observation will be confirmed *in vivo*, then the chromatin might act as a buffer which transiently absorbs torsional stress to keep the chromatin environment comfortable for DNA-tracking complexes [89, 91]. Comparison of the expression profiles of cells - wild type or mutant - for different topoisomerase, revealed that these enzymes play an important role during transcription [69, 92, 93]. According to their capability to cut and reseal one or two DNA strands, topoisomerases are divided broadly into two families: type I enzymes transiently break one DNA strand; type II topoisomerases cleave and rejoin both strands [92]. The ability of the two types of enzyme to efficiently remove both positive and negative supercoiling in eukaryotes reflects a mechanical and functional redundancy between different topoisomerases [69, 92]. Since supercoils generated in front of the transcribing RNA polymerase have a different effect on transcription and reside in a different molecular environment compared to those

generated behind it, different solutions of topological problems and specialized roles of topoisomerases may occur in each circumstance. Indeed, in yeast, positive torsional stress in front of the RNA polymerase I is largely resolved by topoisomerase II (Topo II), while topoisomerase I (Topo I) is responsible for the removal of the negative torsional stress behind the polymerase [94]. Topo II is the main relaxase on chromatin fibers *in vitro* but it binds primarily to the nucleosome-free regions *in vivo* [95, 96]. Notably, under the same experimental conditions, naked DNA was relaxed by Topo I much faster than by Topo II [96]. This finding suggests that Topo I is a more processive and rapid enzyme which probably works near the regions stripped of nucleosomes with a high demand for relaxation, i.e., close to RNA polymerase. In support of this idea, magnetic tweezers experiments also revealed Topo I to be a torque-sensitive enzyme as the mean number of relaxed supercoils increases with the torque stored in the DNA [97].

The complexity of the processes involved in the twist diffusion through the chromatin and their transient nature, as well as the absence of a clear explanation as to how topoisomerases are recruited to active genes have made it very difficult to predict the extent of supercoiling at each particular genomic locus. Our understanding of this multi-factor mechanism is still rudimentary and requires extensive experimental efforts.

As part of this work we are have developed some new insights into understanding of how topoisomerases are recruited to active and inactive genes. See chapter 5 for more details.

## 3.5 Methods to assess the DNA supercoiling

The first techniques to study the torsional state of DNA relied on DNA supercoiling mediated changes in the compaction and the geometry of DNA (Fig. 3.1a) observable by equilibrium and velocity sedimentation, by electron microscopy and by electrophoretic separation [98, 99, 100]. Currently these methods are mostly used for determining supercoiling in populations of circular DNA, i.e. plasmids. These techniques report the average behavior of many DNA molecules and do not characterize the dynamics of structural transitions. During the last one and one-half decades, controlled mechanical manipulation of single DNA molecules or chromatin fibers has been developed to study supercoil-diffusion, the behavior of nucleosome arrays under torsional stress and the active removal of supercoils by topoisomerases [101, 102]. These *in vitro* methods have improved our understanding of DNA mechanics but do not allow monitoring the mechanics and dynamics of the response of DNA to torsional stress in an *in vivo* context.

The degree of supercoiling in intracellular DNA has been estimated most often using a strategy that relies on the binding of various psoralen derivatives to DNA (Fig. 3.2a). The psoralens are cell membrane-permeable molecules with a planar, aromatic structure that allows them to intercalate into B-DNA. The extent of psoralen intercalation is linearly related to the level of negative superhelicity and provides a measure of DNA topology *in vivo* [103, 104]. Such experiments have revealed that although the bulk of genomic DNA is relaxed, supercoiled DNA does exist at a few loci of mammalian cells [105, 106]. In Drosophila polytene chromo-

somes, the pattern of psoralen binding has been used to directly visualize torsionally stressed DNA which appeared to localize at active genes [107]. In a recent modification of the psoralen-based technique, binding of the compound to the yeast genome *in vivo* was examined genome-wide using DNA arrays [108]. It was shown that large chromosomal compartments have different levels of DNA superhelicity but the experiment failed to detect transcription-induced supercoiling, probably due to the high density of genes in yeast and very short linker DNA which together require a method with a higher resolution.

The first direct measure of transcription-generated supercoiling *in vivo* in human cells was made by using a site-specific Cre-recombinase to excise a chromatin fragment upstream of an inducible promoter [31]. Recombinase-mediated circularization of the fragment enabled the trapping of negative supercoils that were diffusing through the chromatin (Fig. 3.2b). This experiment showed that DNA supercoiling dynamically elicits the relaxation potential of topoisomerases [31]. The transmission of negative supercoils upstream of the actively transcribed regions has been demonstrated to occur even on linear DNA *in vitro*, showing that the generation of supercoiling is much faster than the free DNA twist diffusion [28]. In addition, since many promoters are sensitive to DNA supercoiling, indirect studies have been used to monitor the pattern of transcriptional activity to obtain information about DNA topology [76, 109, 110]. DNA topoisomerases also provide a valuable tool to investigate the topology of DNA and could function as *in vivo* probes to measure the level of torsional stress. Given their specialized functions, the mapping of the exact position of topoisomerases along the genome should enable an *in vivo* assessment of

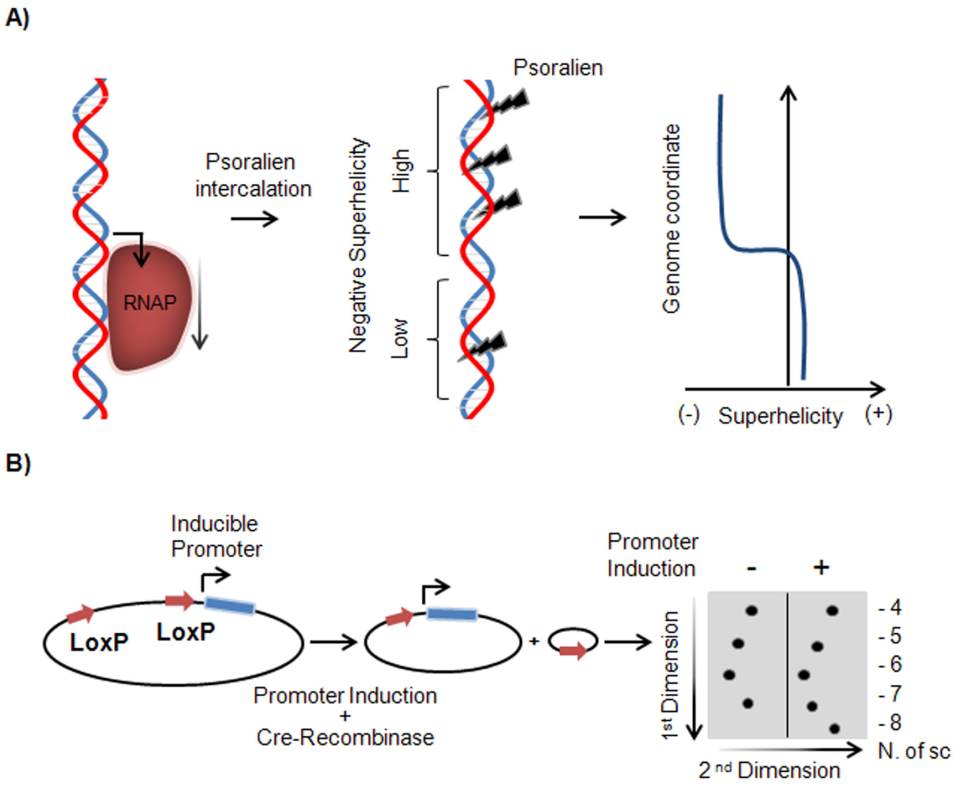


Figure 3.2: Strategies to assess the DNA topology inside of the cells: A) Psoralen intercalates preferentially into undertwisted DNA and, upon exposure to UV-light, crosslinks its strands. DNA supercoiling *in vivo* can be monitored through the extent of photo-crosslinking between different loci in the cell. B) Dynamic torsional stress propagating from an activated promoter between the loxP sites is trapped in the DNA circle excised by Cre-recombinase. Two-dimensional electrophoresis of the circles gives an accurate accounting of DNA supercoiling generated during transcription.

the supercoils distribution [94, 95, 111, 112].

## 3.6 DNA supercoiling in regulatory pathways

In a eukaryotic cell, basal chromatin organization not only prevents access of the RNA polymerase to promoters but also restricts transcription elongation along the DNA. Because of the strong binding energy between nucleosomes and DNA, transcription requires chromatin remodelers to disrupt or to slide nucleosomes, providing a means for transcription regulation. There is substantial evidence from *in vivo* experiments to indicate that nucleosome disruption is needed for proper elongation; importantly, this disruption propagated along the gene faster than the rate of RNA polymerase II translocation [113]. Positive DNA supercoiling promotes unwrapping of DNA from the histones and modifies nucleosome structure *in vitro*; in contrast nucleosomes rapidly form on negatively supercoiled DNA [67]. Consequently, it was suggested that at each round of transcription, the positive supercoiling is pushed ahead of RNA polymerase. Accumulated positive torsional stress induces structural modification of nucleosomes and creates conditions in which polymerase efficiently elongates through the nucleosomal array [90, 114]. Negative stress in the wake of the transcription machinery facilitates rapid re-formation of nucleosomes behind the elongating complex. Thus, by variation in intensity and polarity, supercoiling may directly modulate the conformation of chromatin to satisfy the demand of transcription in real-time (Fig. 3.3a). Indeed, it was shown that treatment of cells with a Topo II inhibitor results in perturbation of chromatin structure,

which seems to indicate that DNA supercoiling mediates chromatin rearrangement [115].

The double helix which is the predominant B-form, could adopt, depending on the sequence composition, a variety of alternative structures [30]. A prerequisite for the formation of these structures is duplex destabilization sponsored by high level of negative supercoiling [116]. In fact, dynamic supercoiling was indirectly measured through the identification of non-B DNA structures in susceptible sequences upstream to active promoters both *in vitro* and *in vivo* [28, 31]. Non-B DNAs bind a diversity of DNA conformation-sensitive proteins some of which have regulatory function, suggesting that these unusual DNA structures are more than mere by-products of genetic activity [30, 117]. Accordingly, *in silico* analyses showed an enrichment of supercoil-sensitive sequences at regulatory loci [118, 119]. To date, the most complete investigation showing the important role of non-B DNA in gene regulation was conducted on the human *c-myc* proto-oncogene. Upstream of the main promoter of MYC it is located a supercoil-sensitive sequence called FUSE. During the transition from the basal level of expression to the full expression in response to activating signals, FUSE starts to melt due to increasing levels of negative supercoiling [29]. Partly melted FUSE binds the transcription activator FUSE-binding protein (FBP), which increases the promoter activity by interacting with the general transcription factor TFIIH and drives the transcription of MYC to peak output. FBP-interacting repressor (FIR) binds FBP and FUSE which is fully melted due to high level of DNA supercoiling. The binding of FIR abolishes the effect of FBP, and the gene transcription is restored to basal levels. Thus, co-

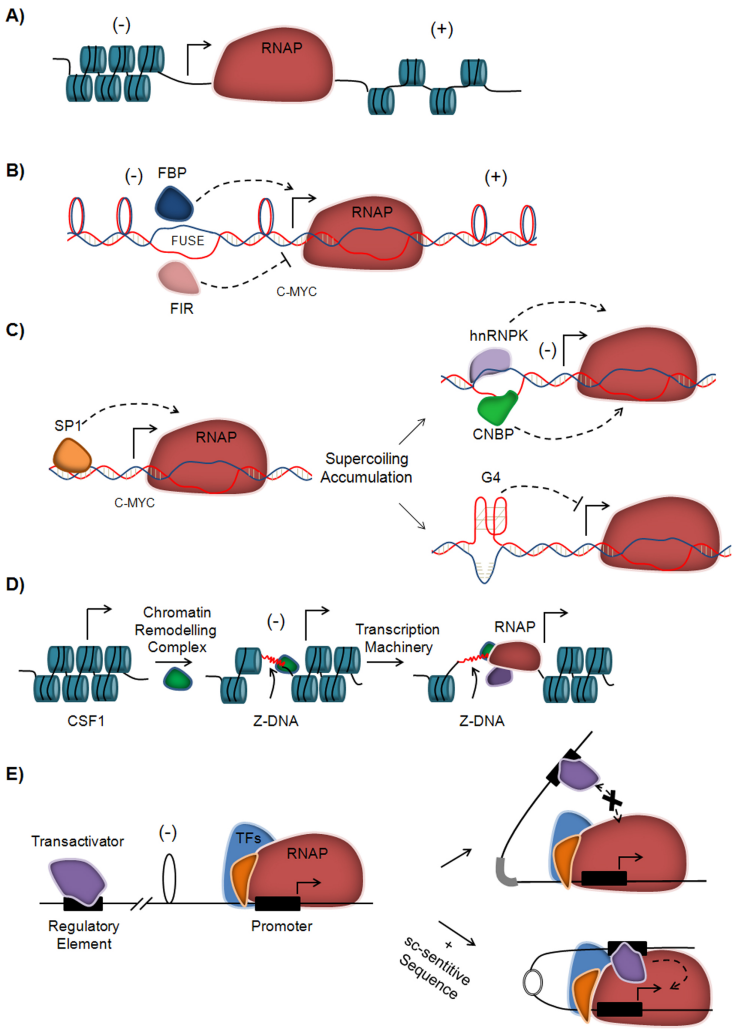


Figure 3.3: Long range regulatory events due to transcription-generated DNA supercoiling: A) Torsional stress modulates the conformation of chromatin, promoting unwrapping of DNA from the histones ahead of RNA polymerase (RNAP) and rewrapping behind it. B) During transcription of *c-myc* gene the melting of the supercoil-sensitive sequence FUSE promotes the recruitment of factors that enhance (FBP) or repress (FIR) the transcription. C) According to the level of torsional stress, the CT-element located upstream of the *c-myc* promoter can flip between different conformations (double-stranded, single-stranded and G-quadruplex/ i-motif) which dictate the binding of specific transcription factors. D) The chromatin remodeling in the promoter of CSF1 favors the formation of Z-DNA which stabilizes the open chromatin structure. (-) means negative supercoils, (+) means positive supercoils. E) Single-stranded structures in supercoiled region provide the flexibility needed to juxtapose distal elements.

operation between supercoil-induced non-B DNA and DNA conformation-sensitive proteins provides a real-time feedback mechanism for controlling gene expression (Fig. 3.3b).

Another important conformationally plastic sequence involved in *c-myc* regulation is the CT-element (also known as NHE III1) located 250 bases upstream of the main promoter [120, 121]. It was observed that this element adopts non-B DNA structures in supercoiled DNA *in vitro* as well as in its endogenous location *in vivo* [122, 123]. In normal B-DNA structure, the CT-element is bound by the transcriptional factor Sp1 which activates transcription. It was suggested, that as a result of supercoil accumulation due to activated transcription, the element flips into the single-stranded conformation and the transcription factors hnRNPK and CNBP bind the purine-rich and pyrimidine-rich strands, respectively, to maintain the active state [121, 124, 125]. Besides the single-stranded conformation, CT-element can adopt stable non-B DNA structures, a G-quadruplex on the purine-rich strand and an i-motif on the pyrimidine-rich strand [126]. These globular structures sequester the transcription factor binding sites and consequently silence transcription. Different sets of binding proteins associate with different conformations of CT-element; consequently, gene specific responses could be achieved using ubiquitous transcriptional factors. Thus the local flipping between different DNA conformations induced by torsional stress plays as a switch in selecting which transcriptional factor to employ according to the physiological demands on the cell (Fig. 3.3c).

One more sequence $1.8\,kb$ upstream of the c-myc promoter has been predicted to assume a left-handed double helical structure called Z-DNA. The region is rec-

ognized *in vitro* by anti-Z-DNA antibodies in permeabilized cells under conditions of active transcription [127]. The function of this sequence in c-myc transcription is currently unknown, although proteins able to specifically interact with Z-DNA have been described [128]. Besides serving as targets for binding, supercoil-induced non-B DNA structures could modify chromatin structure by exclusion of nucleosomes [129, 130, 131]. It was shown that activation of the CSF1 gene by chromatin remodeling activities, results in formation of Z-DNA at the sequence located within the promoter which, in turn, stabilizes the open chromatin structure in the area critical for efficient transcription (Fig. 3.3d). The elastic properties of non-B DNA are different from those of B-DNA. Double helix is a stiff polymer and cells should overcome its rigidity to facilitate DNA-protein-DNA interactions which are playing an important role in many cellular processes [132]. Non-B conformations expose flexible single-stranded segments that together with plectoneme formation may facilitate DNA transaction between flanking sequences (Fig. 3.3e) [133, 134].

MYC deregulation is just one of several crucial hallmarks of cancer. It was suggested that cancer genotypes are set up by eight essential alterations in single cells that dictate malignancy: sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, activating invasion and metastasis, reprogramming of energy metabolism and evading immune destruction [135]. Each of these physiologic changes is manifested by alterations in expression of key genes, with many of them containing supercoil-sensitive sequences in the core or proximal promoter. Given the importance of these genes, including MYC, KRAS, RB1, BCL2, VEGFA, TERT and PDGFA, additional layers

of tight regulation may be imposed at their promoters. The response of CT- and FUSE-like elements to transcription-generated supercoiling reflects the intensity of ongoing transcription, and DNA conformation-sensitive proteins close the real-time feedback loop to provide regulatory adjustment necessary to synchronize the output of gene expression within the population of cells [117, 121].

## 3.7 Summary and Conclusions

In the early days much effort was expended to understand the interplay between the genetic code and chromatin structure: DNA primary structure was found to contain signals that participate in the regulation of DNA metabolism [136, 137]. In the recent years there is a growing body of experimental evidence supporting the idea that DNA mechanics are responsible for a variety of regulatory functions: DNA supercoiling modulates the dynamic rearrangement of chromatin to control the final output of the specific DNA processes [29, 31, 117, 121].

The assembly of multi-protein complexes allows a precise spatio-temporal control of DNA metabolism and particularly of gene expression. By representing the targets of transcriptional factors, cis-regulatory modules provide the essential instructions to coordinate genetic processes. The constellation of factors, both activators and repressors, bound to each module sequence depends on their expression levels. Thus, the variation in the local concentration of transcriptional factors determines the transcriptional outcome, which is a common way to regulate transcription. At the same time, the delay imposed by multiple events necessary to change

the relative concentration of the factors (transcription, translation, protein modification, etc) results in the danger of low synchronization between the physiological requirement and the acute response of important genes such as proto-oncogenes. In contrast, propagation of torsional stress on the DNA is fast and may serve as an efficient long-range signal. The signal could restrict or promote the enrollment of DNA conformation-sensitive proteins at the regulatory module, or could favor the proper arrangement of protein-DNA interaction over long distances. The same regulatory outputs could be reached by adjustment in transcription factor synthesis, but only DNA supercoiling has the capacity to govern the specific transaction moment-to-moment, according to the demands of a DNA-dependent processes.

Our understanding of this phenomenon is still elusive. Although chromatin biology has been gaining much more interest, the associated torsional state of DNA remains neglected since it is less amenable to analysis. Exploring the phenomenon requires the aggressive development of new techniques for measuring of DNA torsional stress with high sequence resolution and preferably at the single-cell level.

# Chapter 4

# Understanding Different Measures for DNA Supercoiling in Microarray Hybridizations

This chapter develops the mathematical framework that enables choosing the correct measure for inference of DNA supercoiling using microarray experiments. All the simulations were performed by me, and they proved very important in our final choice where we decided to drop some of the cross-hybridization experiments that were thought to be good measures of DNA supercoiling earlier.

## 4.1 Overview

Psoralen intercalation has been used as a marker for *in vivo* DNA supercoiling for over three decades now. Combined with microarrays, it becomes an even more useful tool. In section 3.5 we reviewed how various groups have used different measures to estimate supercoiling levels, and make inferences from their observations. Each measure gives us information about DNA supercoiling in slightly different way, and it is important to choose the correct measure to make appropriate inferences. In this chapter, we examine these various measures of DNA supercoiling, in the context of microarray hybridizations, and discusses relations between them.

## 4.2 Psoralen intercalation

As a direct measurement of the supercoiling level is not yet possible for various reasons, we use an indirect method of psroalen intercalation, commonly known as PUVA (or Psoralen + UV Activation).

We know that the intercalation drug psoralen has only a slightly higher preference for intercalation in negatively supercoiled DNA than it has for relaxed DNA. Let us assume that we have a 20% chance of psoralen intercalation in relaxed DNA, which increases to about 35% in negatively supercoiled DNA [43].

To better understand the different measures of supercoiling, let us consider a hypothetical case where we have two distinct sets of experiments, with two distinct levels of negative DNA supercoiling[1] in some region of the DNA.[2]

Note that in principle, we are comparing the same regions of DNA under conditions with different levels of supercoiling, they may come from the same experiment or a different experiment. A schematic diagram is shown in Fig. 4.1. The symbolic reference to DRB here is to give a realistic example.[3] DRB is an elongation inhibition drug, and hence causes the polymerase to stop transcribing. As a result we will loose the transcriptionally generated negative supercoiling but retain the supercoiling due to the inherent chromatin structure, hence two levels of negative supercoiling.

[1] Negative, since we are using psoralen crosslinking as the probe of this supercoiling.

[2] We'll not consider the ideas of positive supercoiling here, as they can be easily extrapolated from this understanding.

[3] As the level of supercoiling in some region upstream to a transcriptionally active gene compared between DRB treated and untreated samples.

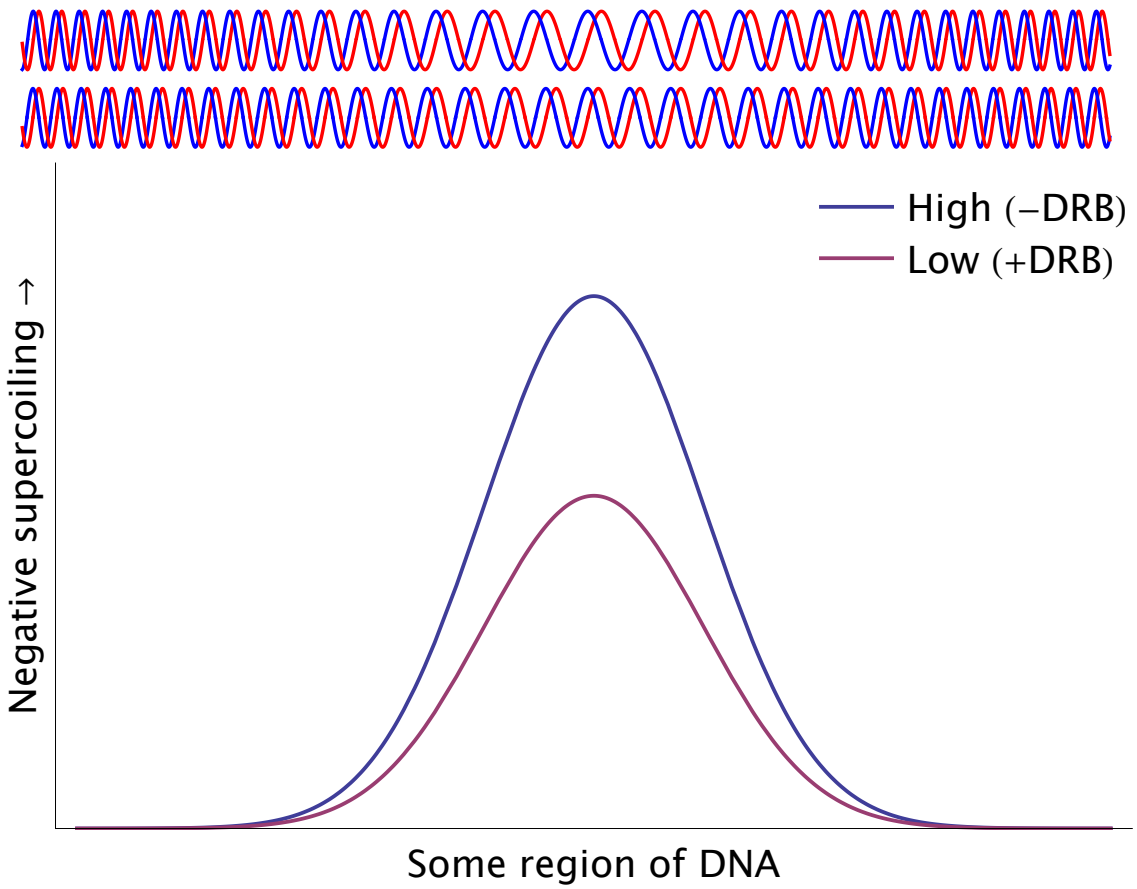


Figure 4.1: Schematic profiles of two different levels of negative supercoiling in the same region of genome between two different experiments.

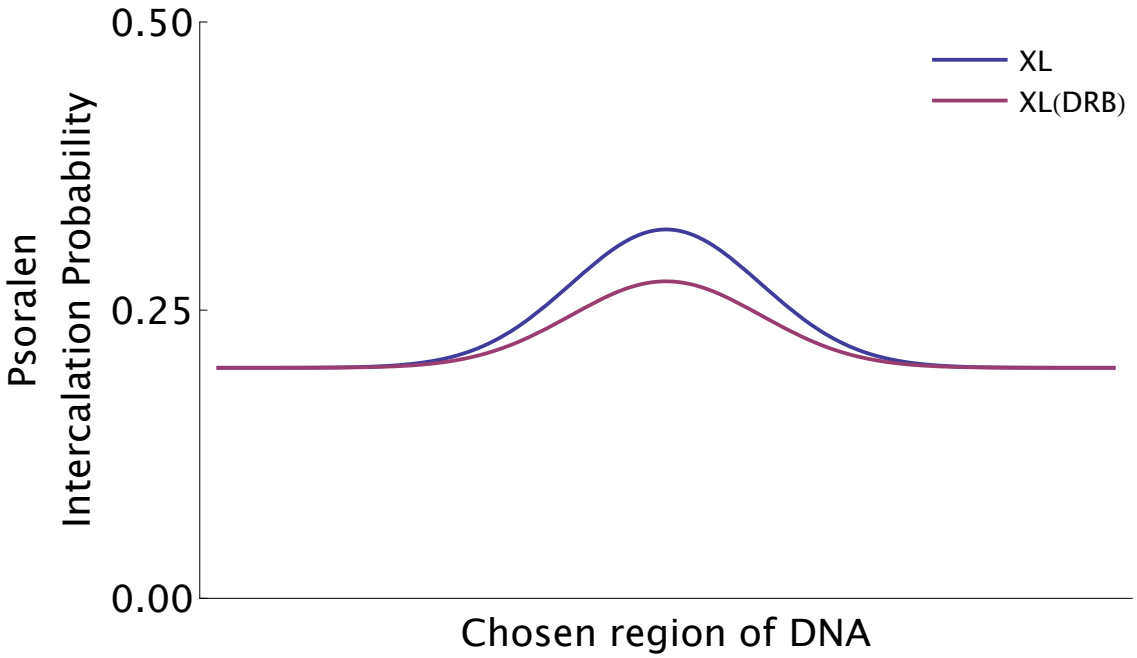


Figure 4.2: Psoralen intercalation probability profiles for the different levels in Fig. 4.1.

The resultant psoralen intercalation probability should look something like Fig. 4.2. Note that as per our previous knowledge, the relaxed regions have a basal level of psoralen intercalation at about 20%, while the regions with increasing levels of negative supercoiling increases the probability of intercalation in the corresponding genomic regions.

## 4.3 Microarray hybridization

After psoralen intercalation and photo-binding, DNA is extracted, denatured and sonicated.[4] Then sonicated DNA is gel-purified to separate the crosslinked ($XL$) and non-crosslinked ($nXL$) DNA. These $XL$ and $nXL$ DNA samples are then hybridized to microarrays. The resultant four hybridizations yield something

[4]For exact details of the experiment refer to section 5.8.

like Fig. 4.3.

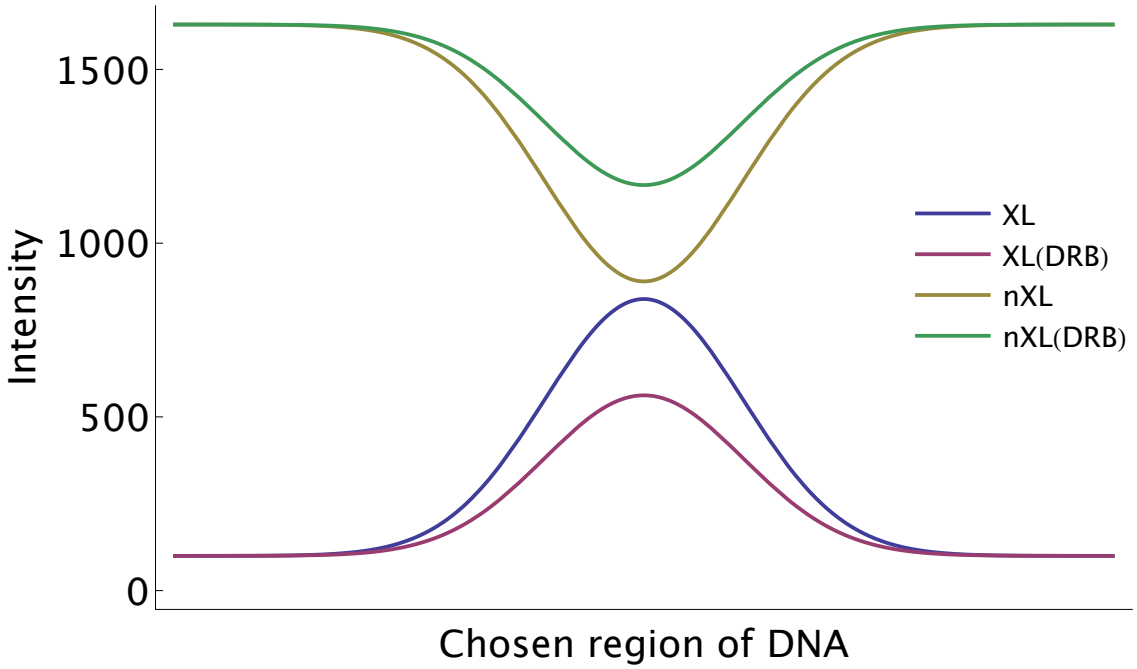


Figure 4.3: Hybridization profiles of the $XL$ and $nXL$ DNA from psoralen intercalation probability profiles in Fig. 4.2.

Here the $nXL$ profiles are obtained using the mass conservation equation on total DNA (i.e. $XL + nXL = const$).[5] In general, the microarray intensities may vary depending on the system, here we are using a maximum intensity of 1024 (i.e. $2^{10}$).

## 4.4 Choosing the correct measures

Our ultimate goal is to make inferences about the relative levels of supercoiling in these two regions.[6] In order to make inferences from these hybridizations,

[5]For more details see equation 4.1 in section 4.6.

[6]These hybridizations give us the level of supercoiling in arbitrary units. We also want to estimate a conversion factor to calibrate the units. See section 4.5 for more details.

we'll have to choose a correct measure of supercoiling. There are several measures possible. Let us start the discussion with primary measures, which are computed from utmost one occurrence of the direct intensities:[7,8]

1. Direct intensity ratios

   - $\frac{XL}{XL(DRB)}$: The ratio of $XL$ intensities from the two hybridizations. A varient of this has been traditionally used as a measure by Richard Sinden.
   - $\frac{nXL}{nXL(DRB)}$: The ratio of $nXL$ intensities from the two hybridizations. Counterpart of the previous one.
   - $\frac{XL}{XL(DRB)} - \frac{nXL}{nXL(DRB)}$: The difference of the two direct intensity ratios.

2. Direct intensity differences

   - $XL - XL(DRB)$: The direct difference between $XL$ intensities from the two hybridizations.
   - $nXL - nXL(DRB)$: The direct difference between $nXL$ intensities from the two hybridizations.
   - $(XL - XL(DRB)) - (nXL - nXL(DRB))$: The difference of two direct intensity differences.[9,10]

[7] For a discussion on secondary measures, see section 4.4.4.

[8] For a discussion on best choice of measures, see section 4.4.5.

[9] Same as: $(XL - nXL) - (XL(DRB) - nXL(DRB))$.

[10] The ratio $\frac{XL - XL(DRB)}{nXL - nXL(DRB)}$ is another possibility but its magnitude should equal one, as the direct intensity differences are expected to be equal to each other in magnitude. See equation 4.1 (section 4.6) for more details.

3. Normalized intensities ratios

   - $\frac{XL}{nXL} - \frac{XL(DRB)}{nXL(DRB)}$: The difference between the normalized $XL$ intensities (normalized by much higher $nXL$ intensities).
   - $\frac{XL}{nXL} / \frac{XL(DRB)}{nXL(DRB)}$: The ratio of the normalized $XL$ intensities (normalized by much higher $nXL$ intensities).[11]

Let us look at these one by one.

### 4.4.1 Ratios of $XL$ and $nXL$ intensities

The ratio of $XL$ and $nXL$ intensities from the two sets of hybridizations, i.e. $\frac{XL}{XL(DRB)}$ and $\frac{nXL}{nXL(DRB)}$, give a conventional measure of relative DNA supercoiling. (The ratio of $XL$ intensities was used by Richard Sinden for his pioneering studies.)

The argument is very simple to understand: If the relative enrichment of $XL$ intensities is higher in the $DRB$ untreated sample, it means that this sample is more negatively supercoiled than the $DRB$ treated sample.[12] By a similar argument for $nXL$ intensity ratios, a higher $nXL$ intensity in untreated sample would mean lower crosslinking and hence lower negative supercoiling. So the $nXL$ ratio profile should look opposite to the $XL$ ratio profile.

Fig. 4.4 shows the cross-hybridization ratios for the hybridizations profiles in Fig. 4.3. As expected the two ratios have opposite nature, although the shapes of the profiles look very different. The $XL$ ratio profile looks considerably flatter

[11] Same as: $\frac{XL}{XL(DRB)} / \frac{nXL}{nXL(DRB)}$.

[12] Higher $XL$ implies higher crosslinking, which implies relatively higher negative supercoiling.

in comparison to the original supercoiling profile (Fig. 4.1) as well as the psoralen intercalation probability profile (Fig. 4.2). (cf. Fig. 4.5 and Fig. 4.7.)[13]

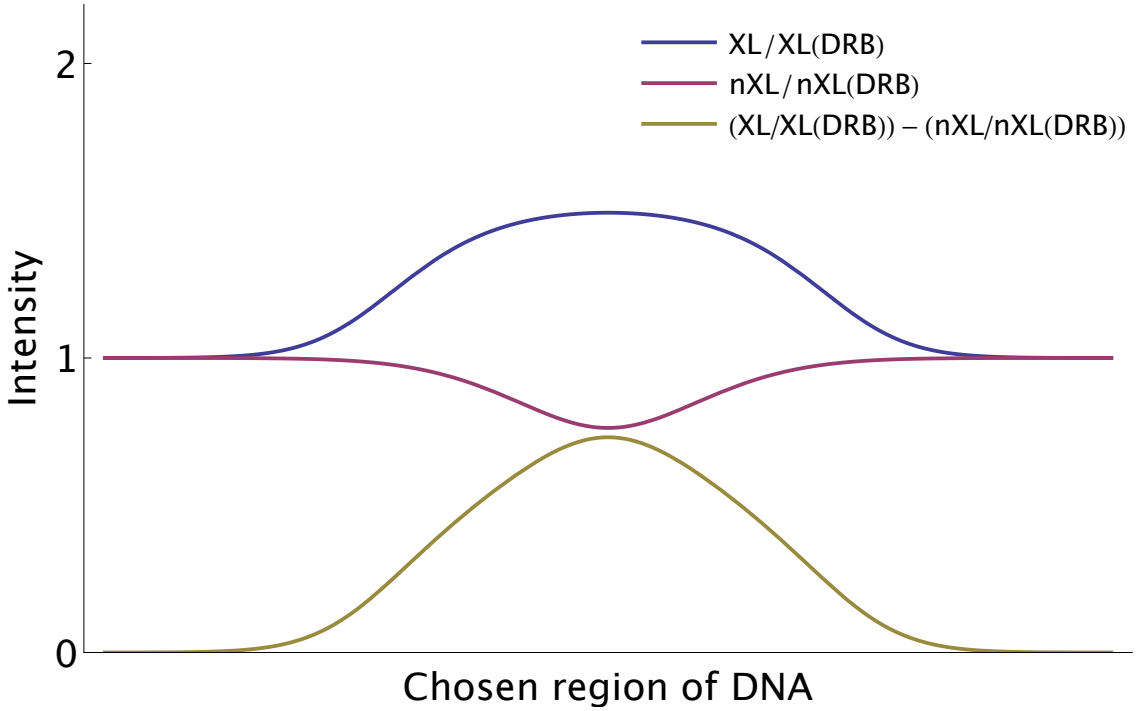


Figure 4.4: Comparing the cross-hybridization ratios (and their difference) for the hybridizations corresponding to Fig. 4.3 profiles. (cf. Fig. 4.5 and Fig. 4.7.)

### 4.4.2 Differences of $XL$ and $nXL$ intensities

The differences of $XL$ and $nXL$ intensities from the two sets of hybridizations, i.e. $XL - XL(DRB)$ and $nXL - nXL(DRB)$, give two direct measures of relative DNA supercoiling. This method is traditionally not used owing to the convention of normalizing to the background in hope of improving the signal to noise ratio (SNR). However, as shown in Fig. 4.5, it turns out that this is the most sensitive method for estimation of supercoiling levels. The issues of SNR can be tackled by taking

[13]See section 4.4.5 for more discussion.

average multiple replicates and by using better smoothing techniques than available before.

Since this is a direct difference, the argument is same as before: If the relative enrichment of $XL$ intensities is higher in the $DRB$ untreated sample, it means that this sample is more negatively supercoiled than the $DRB$ treated sample. Higher $XL$ implies higher crosslinking, which implies relatively higher negative supercoiling. By a similar argument for $nXL$ intensity ratios, a higher $nXL$ intensity in untreated sample would mean lower crosslinking and hence lower negative supercoiling. So the $nXL$ difference profile should look opposite to the $XL$ difference profile.

Fig. 4.5 shows the differences for the hybridizations profiles in Fig. 4.3. As expected the two ratios have opposite nature. The shapes of both the profiles look almost identical to the original supercoiling profile (Fig. 4.1) as well as the psoralen intercalation probability profile (Fig. 4.2). (cf. Fig. 4.4 and Fig. 4.7. See section 4.4.5 for more discussion.)

### 4.4.3 Normalized intensities ratios $\frac{XL}{nXL}$ and $\frac{XL(DRB)}{nXL(DRB)}$

The normalized intensities ($\frac{XL}{nXL}$ and $\frac{XL(DRB)}{nXL(DRB)}$) also give us two measures of DNA supercoiling, i.e. $\frac{XL}{nXL} - \frac{XL(DRB)}{nXL(DRB)}$ and $\frac{XL}{nXL} - \frac{XL(DRB)}{nXL(DRB)}$.

Before discussing the measures, let us have a look at Fig. 4.6 which shows the profiles of the normalized intensities - $\frac{XL}{nXL}$ and $\frac{XL(DRB)}{nXL(DRB)}$.

Since these profiles looks similar to the original supercoiling profile (Fig. 4.1) as well as the psoralen intercalation probability profile (Fig. 4.2), we can use their

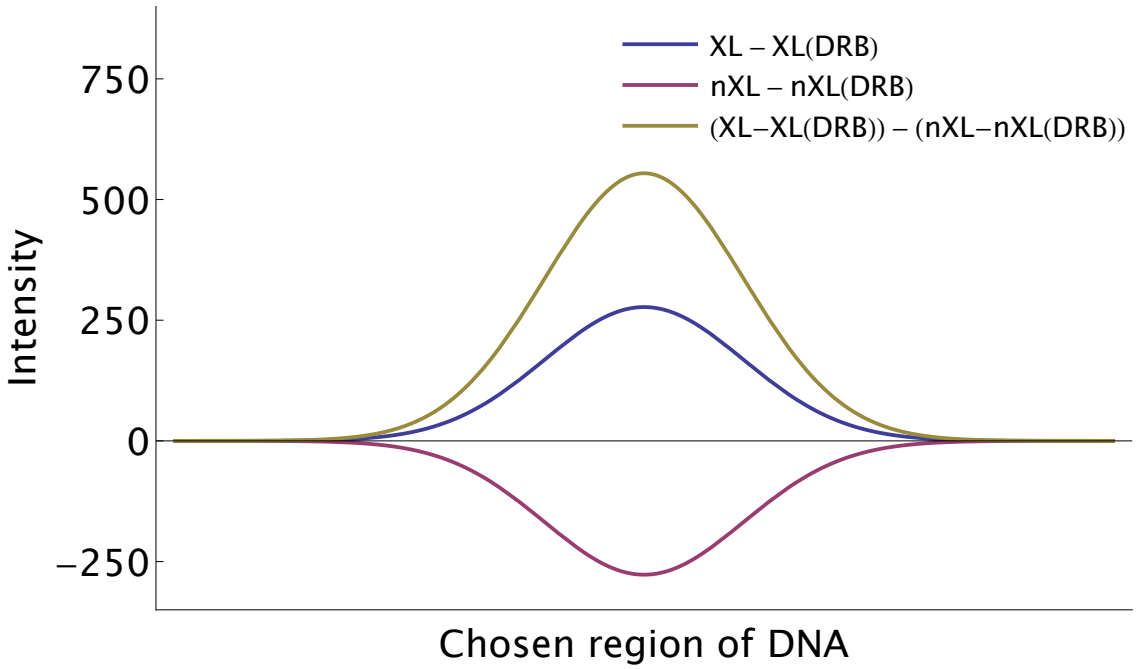


Figure 4.5: Comparing the cross-hybridization differences (and their difference) for the hybridizations corresponding to Fig. 4.3 profiles. (cf. Fig. 4.4 and Fig. 4.7.)

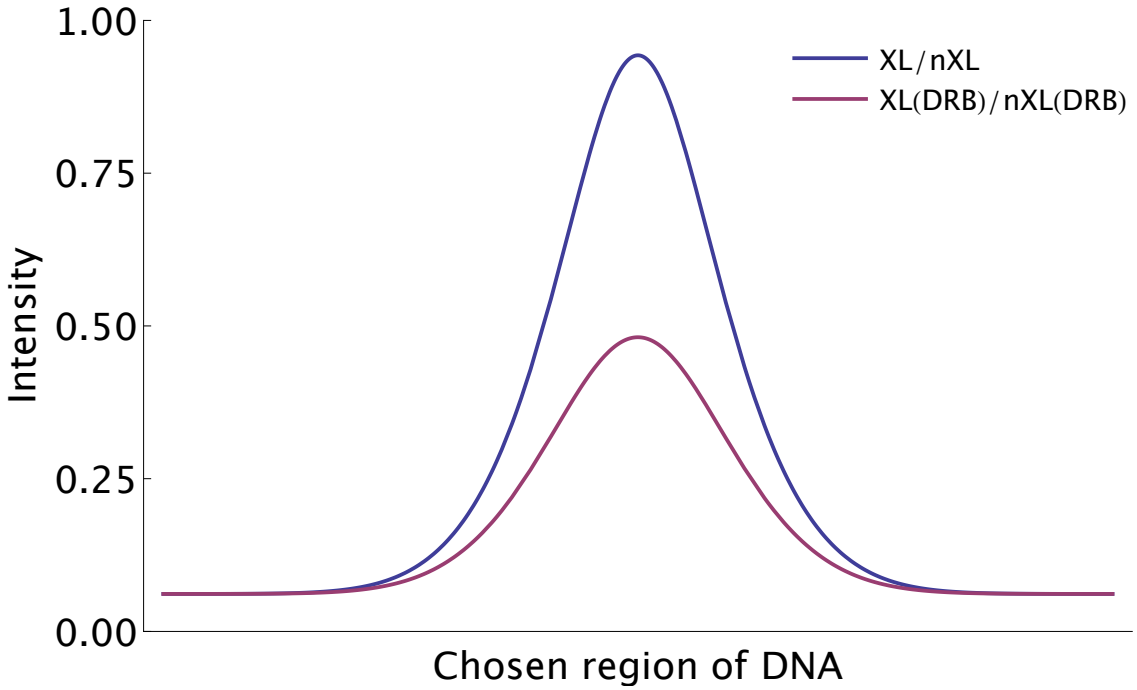


Figure 4.6: Comparing the profiles of the normalized intensities, $\frac{XL}{nXL}$ and $\frac{XL(DRB)}{nXL(DRB)}$. (cf. Fig. 4.1 and Fig. 4.2.)

ratios and differences also as measures of the supercoiling (just like in previous two sections). The justfication is also unchanged, and both measures are expected to be similar to the two supercoiling profiles (Fig. 4.1).

Fig. 4.7 shows the difference and ratio measures of the normalized intensities, along with the logarithm of normalized intensity ratio (on base 2). As expected the two measures have similar profile, although their baseline is different. The difference in baseline is one, and is expected (1 comes from the ratio, and 0 comes from difference). (cf. Fig. 4.4 and Fig. 4.5. See section 4.4.5 for more discussion.)

Fig. 4.7 also shows the logarithm (on base 2) of the ratio profile. Note that it looks similar to the difference profile, but has a larger range. From equation 4.8 (section 4.6), we can see that the log of ratio intensities, i.e. $\log_2 \frac{XL}{nXL} - \log_2 \frac{XL(DRB)}{nXL(DRB)}$ is related to the direct intensity ratio measure in section 4.4.1, and is equal to $\log_2 \frac{XL}{XL(DRB)} - \log_2 \frac{nXL}{nXL(DRB)}$.

### 4.4.4 On secondary measures

So far we have considered the primary measures, i.e. measures which are computed from utmost one occurrence of the direct intensities. It is possible to generate infinitely many new measures by combining the primary measures. In some cases, it might be beneficial to use secondary measures as they can be carefully constructed to increase the dynamic range. However, in most cases they will be redundant. (Note that the log difference plotted in Fig. 4.7 is not a secondary measure, it is just the ratios replotted after taking the logarithm.)

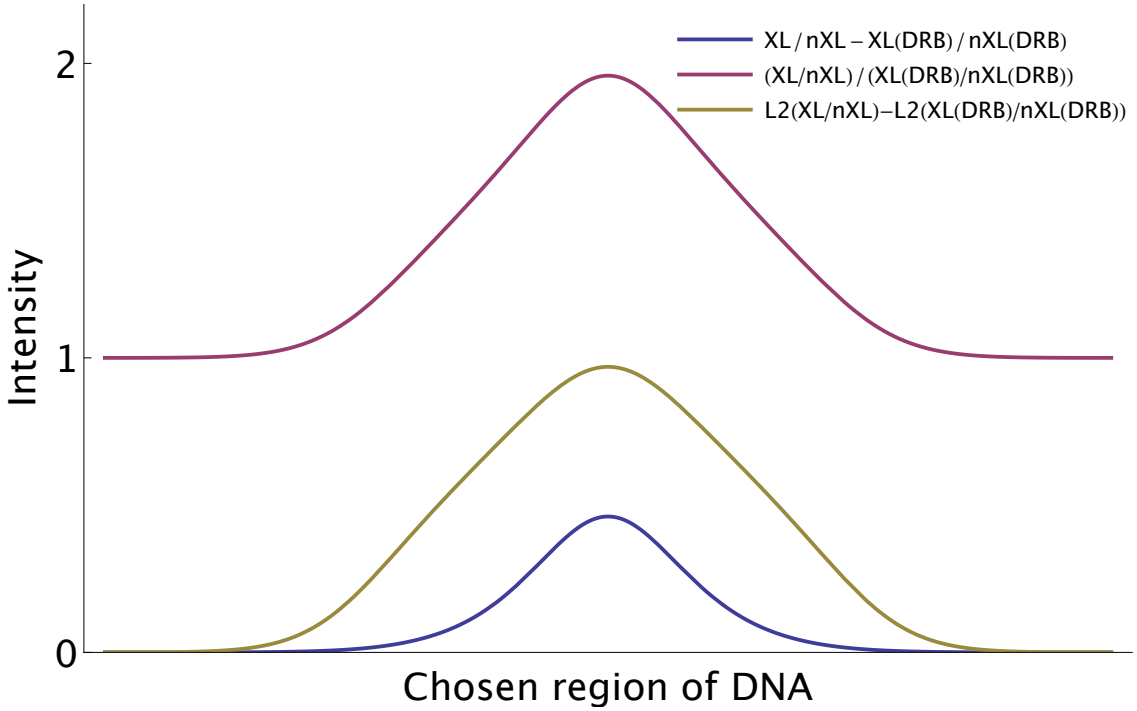


Figure 4.7: Comparing the difference and ratio measures of the normalized intensities. Also note the logarithm of normalized intensity ratio, $L2$ refers to the logarithm taken with base 2. (cf. Fig. 4.4 and Fig. 4.5.)

### 4.4.5 Best choice

Comparing the dynamic ranges and shapes of three pairs of measures in Fig. 4.4, Fig. 4.5 and Fig. 4.7, it seems that the direct intensity differences in Fig. 4.5 give the best measure. We expect this to be true in general because of the wide range of intensities of microarray optical reader.[14] However, owing to the sensitivity of different drugs and specific experimental conditions, it is advisable to test all six measures for specific experimental parameters. (Also see section 4.6.1) In particular, ratios will be better measures when the variance in the direct intensities of hybridization data is very small.

During our experiments the $XL$ and $nXL$ parts of DNA were not loaded in

[14] *NimbleGen* uses a range of $(0 - 32,768)$

the 1 : 3 ratio that they were extracted. Instead, to minimize noise and to compare with previous methods [103], equal quantities of $XL$ and $nXL$ DNA were loaded on the microarray chips. Based on the discussion above, and analysis of our data we find that $\log_2 \frac{XL(DRB)}{nXL(DRB)} - \log_2 \frac{XL}{nXL}$ is a very good measure of supercoiling, provided appropriate sequence dependent corrections are done. (See section 6.5.2 for more details.)

## 4.5 Calibration to estimate $\pm\Delta Lk$

The hybridizations in Fig. 4.3 as well the various measures (in Fig. 4.4, Fig. 4.5 and Fig. 4.7) give us the level of supercoiling in arbitrary units. We also want to estimate a conversion factor to calibrate the units and get the corresponding change in linking number ($\pm\Delta Lk$). We can use the various observations in literature depending on the corresponding measure used there. Section 4.6 gives various relations between these measures.

## 4.6 Mathematical relations among measures

### 4.6.1 Word of caution

This section deals with relations among various measures considered in this report. The first part is dependent on mass conservation and the second part isn't (although the same relations can be derived using mass conservation also).

One must be aware of the following while using these relations to test the data quality: The relations derived using mass conservation may be exact in theory,

but they will be more prone to systematic errors, and very sensitive to sample preparation. The assumption is that total DNA content is same, but this would in principle be very difficult to achieve because DNA might be lost selectively during various steps (e.g. sonication, gel purification). This means that one can still expect the profiles to look similar, but the values may not match exactly.

On the other hand, the relations derived without using mass conservation will be more robust. The fact that this equation holds true in general, means that these relations can be expected to hold at tighter bounds. They should be very suitable for internal controls, and calibration.

### 4.6.2 Mass conservation dependent relations

We are hybridizing the crosslinked ($XL$) and non-crosslinked ($nXL$) DNA to microarrays. Since the total amount of DNA in the DRB treated and untreated cells is the same, we can write the mass conservation equation 4.1.

$$XL + nXL = XL(DRB) + nXL(DRB) \tag{4.1}$$

By simply rearranging the two sides, we get the relation between the direct differences:

$$XL - XL(DRB) = nXL(DRB) - nXL \tag{4.2}$$

Again, by dividing both sides of Eq. 4.1 by $nXL$ or by $nXL(DRB)$, we get equations 4.3 and 4.4.

$$\frac{XL}{nXL} + 1 = \frac{XL(DRB) + nXL(DRB)}{nXL} \tag{4.3}$$

$$\frac{XL + nXL}{nXL(DRB)} = \frac{XL(DRB)}{nXL(DRB)} + 1 \tag{4.4}$$

Now let us multiply equation the left side of 4.3 by the left side of 4.4 and the right side of 4.3 by the right side of 4.4 then substitute (utilizing equation 4.1). This gives us equation 4.5.

$$\left(\frac{XL}{nXL} + 1\right) \frac{1}{nXL(DRB)} = \frac{1}{nXL} \left(\frac{XL(DRB)}{nXL(DRB)} + 1\right) \tag{4.5}$$

Equation 4.5 simplifies to equation 4.6.

$$\left(\frac{XL}{nXL} + 1\right) \frac{nXL}{nXL(DRB)} = \left(\frac{XL(DRB)}{nXL(DRB)} + 1\right) \tag{4.6}$$

Note that ratio $XL/nXL$ represents the relative enrichment of crosslinked DNA. Taking logarithm on both sides, we get equation 4.7.

$$\log_2 \left(\frac{XL}{nXL} + 1\right) + \log_2 \frac{nXL}{nXL(DRB)} = \log_2 \left(\frac{XL(DRB)}{nXL(DRB)} + 1\right) \tag{4.7}$$

Note: Taking the logarithm makes it easy to visualize relative enrichment (or depletion) of intercalation. The unaffected regions fall on the x axis (as they will have a ratio of 1, and $\log_2(1) = 0$). We are taking the logarithm base as 2, as it has been previously demonstrated that the maximum change in the level of supercoiling (*in vivo*) is approximately

two-fold [103, 31]. On $\log_2$ scale, two fold enrichment/depletion comes up in a conveniently readable window of $+1/-1$.

Note: These ratios are calculated for each individual probe of microarray.

### 4.6.3 General relations

Let us consider the following relation:

$$\left(\frac{XL(DRB)}{nXL(DRB)}\right)\left(\frac{nXL}{XL}\right) = \left(\frac{XL(DRB)}{nXL(DRB)}\right)\left(\frac{nXL}{XL}\right)$$

The two sides are identical, and such a relation is always true.[15] Now let us rearrange the right hand side (RHS) numerator slightly:

$$\left(\frac{XL(DRB)}{nXL(DRB)}\right)\left(\frac{nXL}{XL}\right) = \left(\frac{nXL}{nXL(DRB)}\right)\left(\frac{XL(DRB)}{XL}\right)$$

Now, let us take logarithm on both sides (using our convention of base 2):

$$\log_2 \frac{XL(DRB)}{nXL(DRB)} - \log_2 \frac{XL}{nXL} = \log_2 \frac{nXL}{nXL(DRB)} - \log_2 \frac{XL}{XL(DRB)} \qquad (4.8)$$

Equation 4.8 gives a relation between the direct ratios of $XL$ and $nXL$ intensities with the normalized ratios of $XL$ intensities. We could also use it as the verification rule and internal control, by means of four separate hybridizations. (Also, see section 4.6.1, c.f. Fig. 4.4 and 4.7.)

[15] We could have invoked the mass conservation law (eq. 4.1) to derive this, but it's not necessary.

# Chapter 5

# Differential tuning of dynamic supercoiling by topoisomerases I and II across the genome

This chapter serves as the main results and discussion section of the first part of this dissertation, i.e. role of DNA supercoiling in gene regulation and control of biological noise. The chapter has been submitted for publication [138]. I am co-first-author on this paper along with Dr. Fedor Kouzine. Other authors were Dr. Laura Baranello, Dr. Khadija Ben-Aissa and Dr. David L. Levens. F.K., K.B. and D.L. designed research, F.K. and L.B. performed all the wet-lab experiments, A.G. developed the mathematical frame work (see chapter 4 and 6) and performed all the bio-informatics analysis, A.G., F.K., L.B. and D.L. analyzed data and wrote the paper.

## 5.1 Overview

Dynamic interplay between DNA, chromatin and the transcription machinery is fundamental for the proper regulation of gene expression. The mechanical forces imparted onto the template and its embracing chromatin have the potential to modify directly the topological state and structure of the DNA and the arrangement of nucleosomes. Though this is a consequence of gene activity, these modifications are

increasingly recognized as a means to provide real-time feedback to the transcription apparatus and to modify gene expression. To disentangle the mostly theoretical connection between transcription and DNA dynamics, we charted an ENCODE map of transcription-generated dynamic supercoiling in human cell line using psoralen photobinding to probe DNA topology *in vivo*. Dynamic supercoils reside within $\sim 2\,kb$ of transcription start sites of almost all active genes. This torsional stress is handled differently between low and high output promoters as shown by experiments using inhibitors of RNA polymerase and topoisomerases, as well as by chromatin immunoprecipation studies of topoisomerase I and II. High output promoters recruit topoisomerase II to upstream regions whereas low levels of dynamic supercoiling are managed by topoisomerase I. The functional coupling between transcription and DNA topology emphasizes the importance of DNA supercoiling for gene regulation.

## 5.2 Introduction

Chromatin is a highly dynamic structure; the panel of bound regulatory proteins, nucleosome composition, DNA and histone modifications, linker histones etc. all may vary temporally across a gene and its *cis*-regulatory elements. In addition, the structure and topology of DNA may change according to the nature and intensity of nearby genetic transactions [19]. Translocation of RNA-polymerases along the double-helix necessary creates torsional stress, which results in strong changes in the topological state of DNA known as supercoiling [48]. These changes have the potential to facilitate or impede virtually all DNA-dependent processes or to serve

as regulatory signals detected by molecular partners [55]. Thus, beyond serving as a passive repository of information, DNA could actively participate in the real-time regulation of genetic processes [89]. Most of what we know about the dynamics of transcription derives from investigations focused on the roles played by the proteins. Though these protein-centric experiments have been crucial in defining the factors involved in transcription, they have tended to neglect a potential role for DNA structure and topology in gene regulation.

Transcription and DNA topology are inexorably linked. As DNA is screwed through the transcription machinery, it follows a helical path dynamically driving positive supercoils ahead and trailing negative supercoils behind the translocating RNA polymerase [70]. Negative supercoiling untwists while positive supercoils overtwist DNA. If translocation proceeds without pause, then RNA polymerase would generate $\sim 7$ supercoils per second [23], and unless dissipated this torsional stress would rise to enormous levels disruptive to all genetic processes [19, 75]. Positive and negative supercoils are relieved by DNA topoisomerases that transiently break and then rejoin the DNA backbone [139]. Depending on the intensity of ongoing transcription and the disposition of topoisomerases, the amount of supercoiling generated locally might exceed the relaxation capacity of nearby DNA topoisomerases leaving the residual DNA torsional stress to propagate through the embracing chromatin [140]. This stress might influence the binding of regulatory proteins to the DNA, change the mobility of nucleosomes and reorganize the architecture of chromatin fiber [55]. Supercoiling may also drive duplex B-DNA into single-stranded or other non-B DNA conformations [121]. Such changes in DNA structure may alter

the ability of DNA and chromatin to loop and twist and so modify the function of enhancers and other *cis*-control elements [30]. Non-B DNA segments may enable the binding of proteins specific for alternative structures, and because non-B DNA is incompatible with nucleosome binding, these structures may help to sustain nucleosome free regions [130]. Since the magnitude and distribution of supercoiling forces throughout the genome are not known, the extent to which any or all of these potential regulatory mechanisms are realized *in vivo* has been a matter for speculation.

The accumulation and propagation of torsional stress along a DNA fiber should depend on many factors including the rate of transcriptional elongation, the length of the transcribed unit, and the spatial arrangement of promoters (for example, divergent promoters, a common motif in mammalian cells, would generate mutually reinforcing upstream negative supercoils) [85, 141, 142]. How torsional stress is transmitted through DNA will depend on the topological domains formed by protein-DNA interactions or by the anchoring of DNA to immobile nuclear structures [88]. Such domains may concentrate or exclude supercoils from selected zones within the chromatin fiber. The binding of other proteins, nucleosome positioning, and histone modifications might all influence the transmission of torsional stress or the activity of topoisomerases. Fundamental to elucidate the role and the control of torsional stress in gene regulation is the understanding of its disposition within chromosomes.

Although in bacteria, chromosomes are organized into domains, in which DNA supercoiling is maintained within precise limits by the balance of supercoil-

modulating activities, in metazoans, whether DNA torsional tension is regulated or is regulatory is less clear and remains controversial [60, 143]. Recent studies in the yeast and fly have provided a coarse-grain view of the distribution of torsional stress along chromosomes, but low resolution has hampered the analysis of the factors that govern the generation, relaxation, and transmission of DNA supercoiling at individual genes *in vivo* [107, 108]. In mammalian cells supercoiling has been studied at only a handful of genes [106, 105]. Torsional stress has been measured by monitoring the supercoiling of plasmids/episomes recovered directly from cells or after excision from chromosomes, and has been inferred from supercoil-dependent structural transitions in DNA or from the activity of supercoil-dependent recombinases [88, 31, 76, 144]. The degree of crosslinking of the intercataling agent psoralen has also been exploited to measure torsional stress; intercalators in general insert between the bases of underwound DNA more easily than of more tightly wound helices where the bases are squeezed together [103]. The low resolution or low throughput of these methods have provided a limited view of the interplay between the factors determining the generation, relaxation, and transmission of DNA supercoiling *in vivo*.

To address this issue, genomic oligonucleotide microarrays were probed with psoralen photo-crosslinked, labeled DNA to chart a genome-scale map of transcription generated dynamic supercoiling *in vivo*. These studies demonstrate that transcription-generated negative supercoiling near the promoters is a common characteristic of virtually every transcribed gene and is transmitted through chromatin as far as 2 kb upstream from the transcription start site. High levels of transcrip-

tion support higher levels of supercoiling that are balanced by the recruitment of topoisomerases. Both topoisomerase I (Topo I) and topoisomerase II (Topo II ) are differentially recruited and distinctly deployed illustrating the interconnection between DNA supercoiling and gene regulation.

## 5.3 Overview of the approach

Our approach exploits the well-established preferential binding of psoralen to supercoiled DNA both *in vitro* and *in vivo* (Fig. 5.1). This cell membrane-permeable molecule intercalates preferentially into undertwisted double helix and crosslinks the complementary DNA strands upon exposure to UV- light [145]. In addition to supercoil dependence, psoralen intercalation is favored by high A-T content and is sterically inhibited by nucleosomes or other DNA-protein interactions [108]. The influence of DNA sequence and chromatin affects the proper estimation of supercoiling. To quantify transcription-generated torsional stress—dynamic supercoiling—, the extent of *in vivo* psoralen intercalation was compared between cells sustaining normal transcription and cells in which transcription was specifically inhibited just prior to crosslinking. The comparison of these datasets intrinsically normalized for the effects of sequence and for the perdurance of DNA-protein interaction (such as nucleosomes), represents the effect of dynamic supercoiling on psoralen crosslinking.

To measure the extent of crosslinking throughout the genome, the human B-cells Raji were treated with psoralen and UV-light. To minimize the influence of DNA replication or mitosis on DNA topology only cells in G1 were assayed

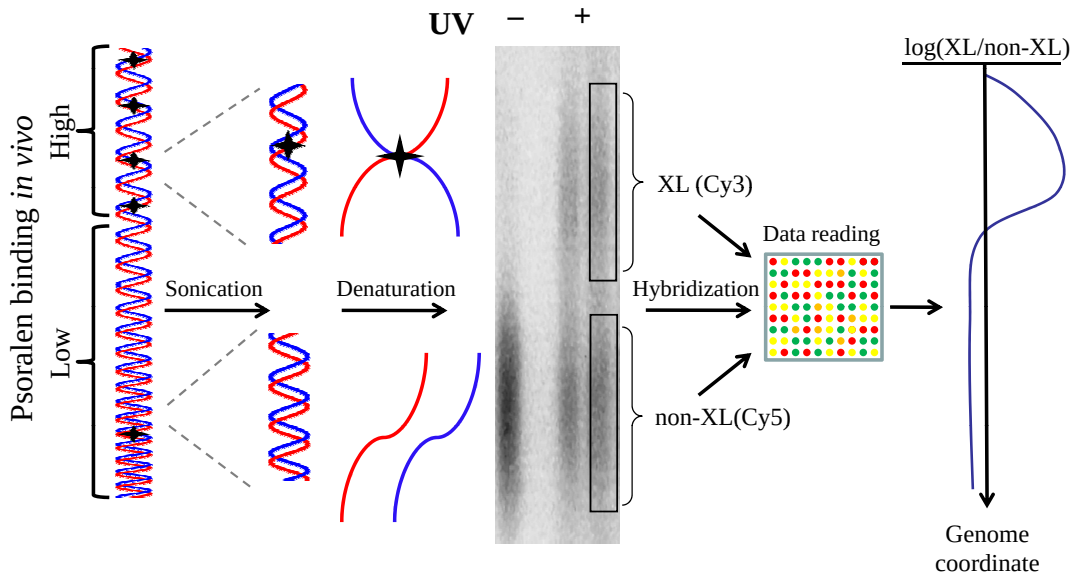


Figure 5.1: Overview of the approach: a scheme for using DNA crosslinking mediated by psoralen photobinding as a genome-probe for DNA supercoiling *in vivo*. Treatment of cells with psoralen followed by UV irradiation produces DNA interstrand crosslinks. Thermal denaturation of genomic DNA fragments results in the formation of two fractions (left). The highly cross-linked fraction (XL) migrates slowly in denaturation gels, while the uncross-linked (non-XL) population is composed of rapidly migrating single-stranded DNA (center). After electrophoretic separation these fractions are purified, fluorochrome labeled and hybridized with densely tiled oligonucleotide arrays (right). The genomic distribution of the ratio of cross-linked and uncross-linked DNA (log 2 scale being 0 at the global mean) represents the efficiency of psoralen intercalation.

[31, 146]. Genomic DNA was recovered, sonicated, denatured and electrophoretically fractionated to resolve the slowly migrating crosslinked population from the faster mobility un-crosslinked one. The crosslinked fraction is enriched with DNA negatively supercoiled *in vivo* at the moment of UV-irradiation (Fig. 5.1). The separated DNA fractions were labeled with Cy5 or Cy3 and hybridized to a high-density oligonucleotide DNA microarrays spanning ENCODE regions [147]. The log ratio (crosslinked/un-crosslinked) of the resulting fluorescent signals which is named CrossLinking level (CL) provides a continuous picture of psoralen intercalation as a function of the genome coordinate. Exemplary results from two genes are shown as a curve smoothed by sliding window averaging (Fig. 5.2a).

In agreement with expectation, promoter areas of gene show markedly different CLs compared with intergenic regions reflecting their enrichment in CpG islands, specialized chromatin structures, and DNA topology. Because physiologically achievable levels of negative supercoiling increase the probability of psoralen intercalation only about two fold relative to relaxed DNA [103], the resulting signal to noise ratio necessitates the use of experimental replicates in order to achieve statistical significance. As described in the methods sections 5.8 and 6.5, three biological replicates were used to generate the CL and other maps analyzed here. The resulting data were averaged across the replicas and the high frequency noise was filtered by Fourier convolution smoothing [148] (supercoiling levels would be expected to fluctuate on the scale of the torsional ($\sim$300 bp) and bending ($\sim$150 bp) persistence lengths of DNA and not base pair-to base pair). To observe the distribution of transcription-generated supercoiling, the CL maps of untreated cells

and DRB (an inhibitor of transcription elongation)-treated cells were compared. Computational subtraction allowed the separation of the effects of chromatin and sequence to reveal directly the component of crosslinking reflecting torsional stress. Similarly, to reveal the dynamic character of DNA supercoiling and to examine its regulation, untreated versus topoisomerase inhibitor treated cells were compared. Because different inhibitors act at different points in the topoisomerase reaction cycle, the changes in DNA topology subsequent to treatment would reflect their modes of action [149].

To relate supercoiling with transcription, nuclear RNA was hybridized with the same microarrays used to assess supercoiling. To correlate the pattern of psoralen intercalation with gene expression, genes were ranked according to their RNA output: from the highest (100%) to the lowest (0%) abundance. We classify genes in three categories: low (0-40%), medium (40-60%), high (80-100%) (see chapter 6). Because closely situated transcriptional start sites (TSSs), especially divergent promoters, could complicate the analysis of DNA supercoiling [31], the set of the ENCODE genes was filtered to remove from the analysis promoters located in close proximity to each other (see section 6.5 for definitions).

## 5.4 Dynamic supercoils upstream of promoters

The dynamic range of gene expression is very large, so mechanistic and structural differences between genes at the extremes of this range might obscure visualization of the basic elastic response of chromatin to applied torque. Therefore,

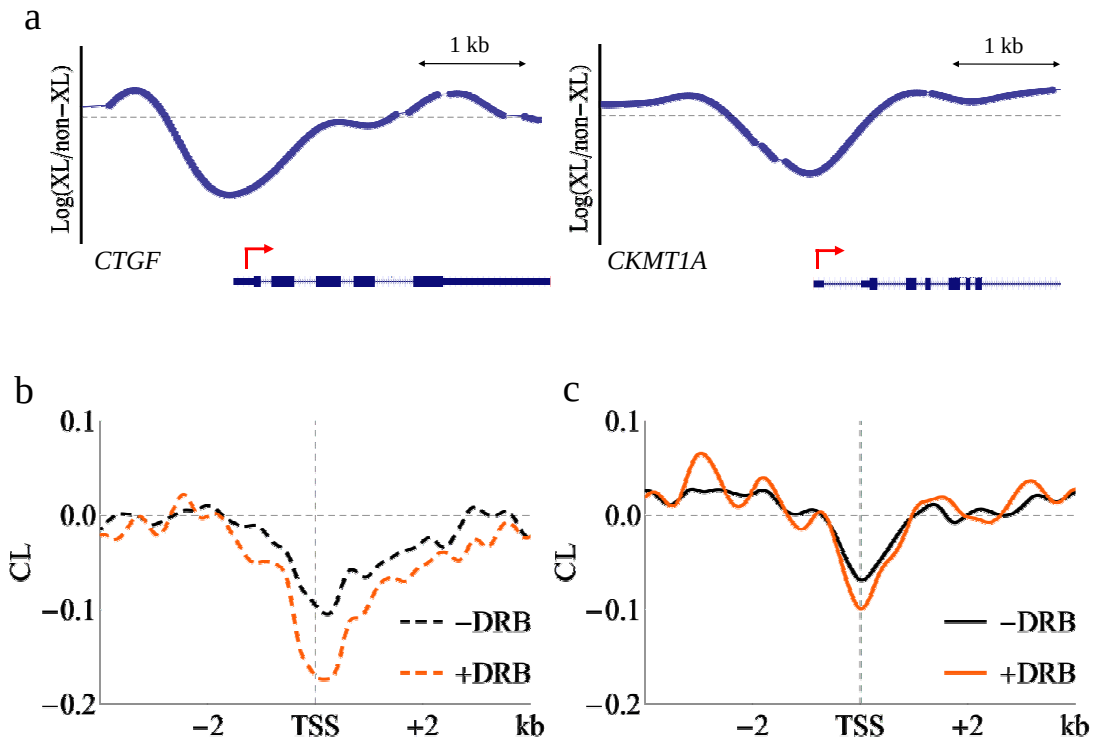


Figure 5.2: Topography of psoralen crosslinking around transcription start sites (TSSs). (a) Representative examples of the psoralen crosslinking map shows peculiarities near TSSs. Composite analysis of psoralen crosslinking levels (CL) near the transcription start sites of medium- (b) and low- (c) expressed ENCODE genes before and after treatment of cells with DRB.

first we compared the smoothed CL profiles of 8 kb windows surrounding TSSs of low (0-20%) and medium expressed (40-60%) genes to see if we could detect modest differences in torsional stress. Meta-analysis of the data for both sets of promoters revealed troughs of overall CL (Fig. 5.2c) at TSSs as expected because these sites reside in psoralen unfriendly CpG islands, heavily laden with transcription and chromatin complexes. The CL profiles were generated also for the cells treated with DRB. DRB specifically inhibits CDK9-mediated phosphorylation of the RNA polymerase II to inhibit transcription elongation [23]. After a short interval of DRB-treatment, the dynamic supercoiling decays as result of topoisomerase activity or diffusion of the torsional stress away from the active promoters. In contrast, DRB should have only small effect on the low-expressed or silent genes with little resident torsional stress. Consequently, after DRB-treatment, the CL-profile reflects only sequence and chromatin, but not dynamic supercoiling. Indeed, the differences in the CL between DRB-treated and untreated cells is maximal near transcriptional start site and gradually declines up to $\sim$2 kb upstream for medium expressed genes (Fig. 5.2b). DRB-inhibition of transcription has little effect, if any, on CL at the TSSs of low expressed genes just as predicted (Fig. 5.2c).

To generate a metric of dynamic supercoiling, we define a parameter called CrossLinking Difference (CLD) from the simple equation: $CLD = CL(+DRB) - CL(-DRB)$ . Thus, CLD is the computational difference between CL values derived from DRB-treated and DRB-untreated cells and is a measure of transcription-dependent psoralen crosslinking. This difference should cancel effects due to DNA-protein interactions and sequence composition. To explore the relationship between

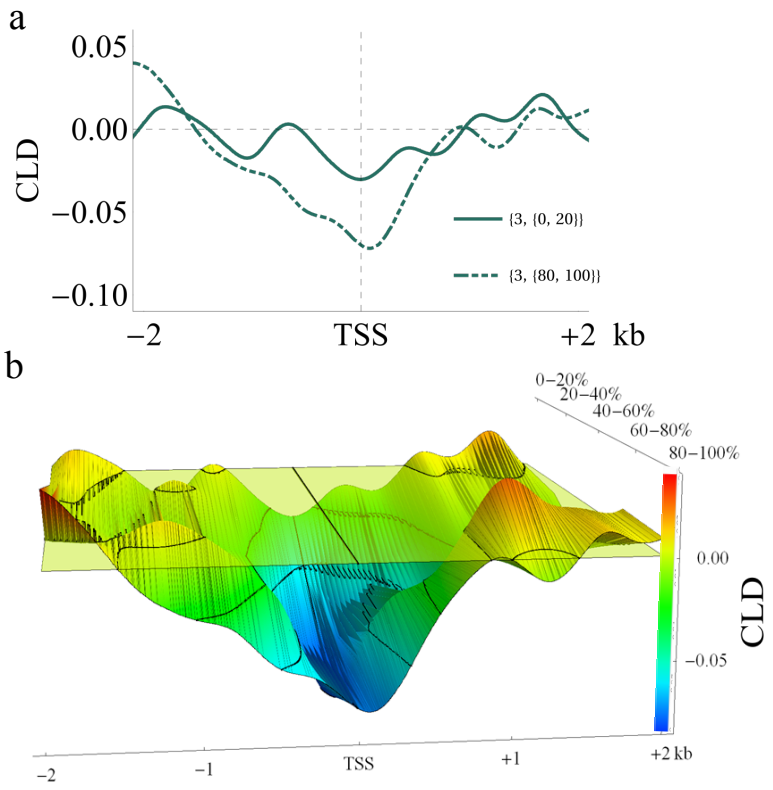


Figure 5.3: DNA topology around TSS as a function of gene expression. (a) Transcription generated supercoils are transmitted up to 2 kb from TSSs. The CrossLinking Difference (CLD) curves of low- and high- expressed genes in a 4 kb window centered at the TSS. Negative CLD values reflect a higher propensity of psoralen to intercalate into the DNA due to transcription-generated supercoiling. (b) 3-D representation of CLD profiles averaged according to the level of gene expression in a 4 kb window surrounding the TSS.

CLD and the transcriptional activity, we compared the average experimental CLD profiles of low and high active genes in 4 kb window surrounding the TSS (Fig. 5.3a). This comparison reveals that during transcription, negative supercoiling is transmitted upstream decaying to baseline about 2 kb from the TSS. As expected from the twin-supercoiled-domain-model, the CLD is diminished within gene bodies where RNA polymerases constitute a moving node between positive and negative supercoils, and because on genes with multiple elongating RNA polymerases, the positive and negative supercoils annihilate each other in inter-RNA polymerase region [48]. At the same time, our analysis was restricted to upstream regions in order to mitigate potential confounding complications. Passage of elongation complexes through gene bodies is necessarily associated with a moving boundary between positive and negative torsional stress, and with the dynamic disruption and subsequent reassembly of nucleosomes and with histone modificatons. Each of these processes affected by DRB treatment has the potential to dramatically alter the type of torsional stress (positive or negative), as well as the degree and distribution of downstream psoralen intercalation and crosslinking. Upstream regions in contrast would be anticipated to subjected mainly to negative supercoiling forces emanating from downstream sources (although the intensity of these forces might fluctuate).

The CLD value reflects a differential psoralen crosslinking derived from the presence of dynamic DNA supercoiling at the upstream promoter region as result of active transcription. Consequently it allows us to make an estimation of supercoiling density. Based on the calibration of psoralen intercalation into the plasmids with defined topology, DNA supercoiling density due to transcription could reach the

$\sigma = -0.08$ near the TSS and then gradually decline into the upstream region.

## 5.5 Parameters controlling the level of dynamic supercoiling

If the twin-domain-model adequately describes the mechanics of transcription, then three major factors define the DNA topology of regions upstream of promoters: 1) the rate of supercoil generation by RNA polymerase; 2) how efficiently torsional stress is transported to remote chromatin locations by twist-diffusion or en-bloc rotation of chromatin segments; and 3) the rate of supercoil removal by topoisomerases [55]. Additional experiments and analyses were conducted to examine the contributions of these parameters to the level of upstream supercoiling.

Transcription-generated supercoiling should increase as gene expression increases unless topoisomerase activity increases in parallel; at steady state, transcriptionally generated torsional stress will be balanced by topoisomerase activity. If torsional stress is freely transmitted through DNA fibers, then increased supercoiling near transcription start sites will be propagated to more upstream regions, unless there are barriers to twist/writhe diffusion. To examine DNA supercoiling as a function of gene expression, CLD signal were averaged for 0-20%, 20-40%, 40-60%, 60-80% and 80-100% expressed genes. CLD strongly correlated with transcriptional activity as predicted by the twin-domain model [48]. Low expressed genes have only a small perturbation of DNA topology in close proximity to the TSS, but as RNA production intensified, the CLD signal spread from 1kb to 2 kb upstream TSS (Fig. 5.3b). The dependence of CLD on gene expression is linear from 0 to 60%

of maximal expression. As RNA production further intensified, DNA supercoiling ceased spreading and declined near TSSs. This result suggests that above the 60th percentile, special mechanisms are marshaled to contend with the highest levels of torsional stress.

The meta-analysis employed above reveals the overall trends in large set of the genes but does not indicate the spectrum of supercoil changes across the ENCODE set of genes. To confirm the unexpected decrease in dynamic supercoiling at the highest levels of expression, genes were ranked in order of promoter output and graphed against the window average of their TSS to -800 bp associated CLDs to display the relationship between expression and torsion (Fig. 5.4a). Indeed, higher output promoters were less supercoiled than medium expressed genes (Fig. 5.4b). In contrast, the CLDs of the -4000 to -4800 bp region were not related to the expression levels. Thus, upstream negative supercoiling is a general feature of transcribed promoters, but the plateau associated with the most active genes indicates that distinct mechanism of DNA relaxation is required to support maximal promoter output.

One simple mechanism to reduce supercoiling near the high output promoters would be to recruit Topo I and/or Topo II more effectively. For validation of this hypothesis, the upstream promoter areas of genes in selected regions across the expression spectrum (0-5%; 55-60%, and 95-100%) were analyzed by chromatin immunoprecipation and qPCR in order to measure their relative levels of associated topoisomerases I and II (Fig. 5.4c). To enhance the detection of catalytically active enzymes, we specifically trapped Topo I and Topo II covalently bound to the

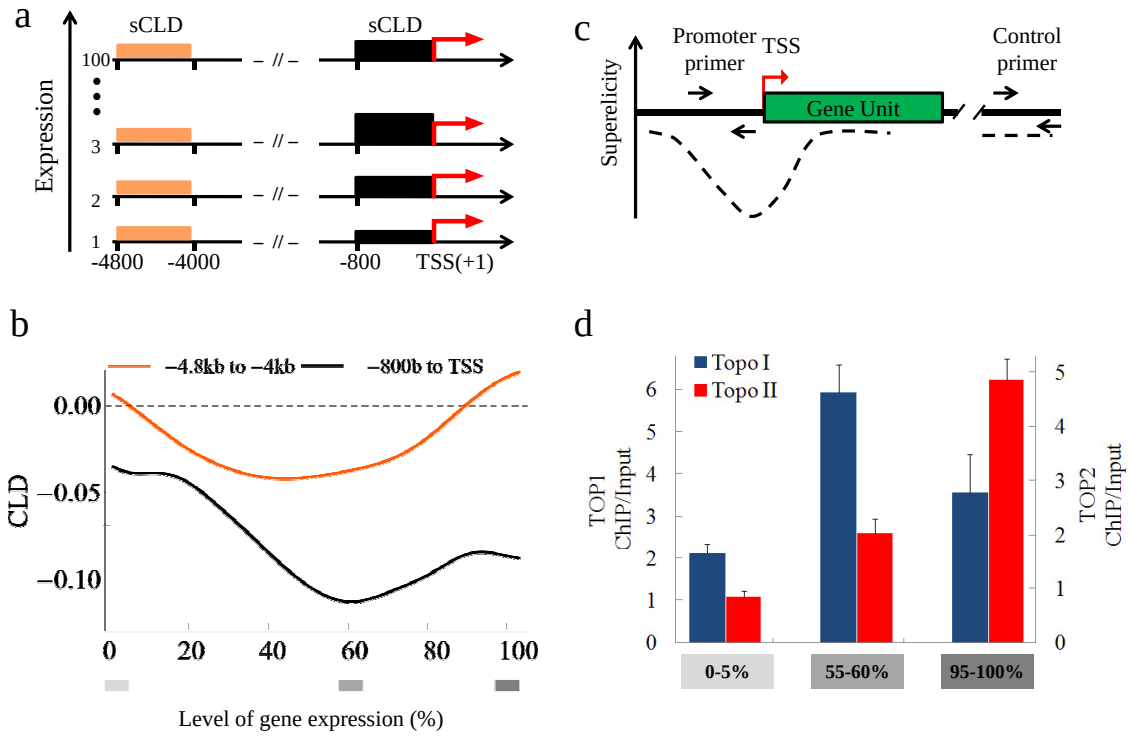


Figure 5.4: Differential patterns of supercoils generation and topoisomerases activities for low-to-medium versus high transcribed genes. (a) Schematic representation describing the calculation used to determine the relationship between expression and DNA topology. (b) The CLD signal of upstream promoters regions was averaged over 800 bp for each single gene and plotted against the level of gene expression (black curve). Smoothing of the curve was done by sliding window average. The CLD signal between -4800 bp and -4000bp (red curve) was graphed for comparison. Gray-scale bars indicate gene expression-ranges from which genes were chosen for ChIP analysis (below). (c) Chromatin from CPT or $\beta$-LAP treated cells was incubated respectively with anti-Topo I or -Topo II antibodies, and the recovered DNAs were analyzed by qPCR using sets of primers spanning promoters versus non-transcribed regions. (d) Average relative enrichment of the genes representing different expression levels analyzed by ChIP for Topo I (blue bar) or Topo II (red bar). Relative enrichment for topoisomerases I and I for each individual gene is shown in Fig. 5.9.

DNA by very brief treatments with camptothecin (CPT) and $\beta$-lapachone ($\beta$-LAP), respectively. CPT is a highly selective drug that inhibits strand religation during Topo I catalytic cycle while $\beta$-LAP poisons Topo II during the formation of the DNA cleavage complex and inhibits Topo I prior to the strand cleavage [150, 151, 152]. The low transcribed genes (0-5%) show very small enrichment relative to the control region for both topoisomerases (Fig. 5.4d). The recruitment of Topo II was dramatically enhanced at the highly active genes. In contrast Topo I was most efficiently recruited to the promoter proximal regions of medium expressed genes (Fig. 5.4d). These results suggest that Topo I and Topo II are differentially recruited to promoters according to their output levels.

Dynamic supercoiling appears to be balanced by topoisomerase action according to their specific distribution and kinetics. To confirm the relationship between DNA-relaxing activities of Topo I and Topo II with gene expression, the CLDs in promoter regions were compared between cells treated with or without topoisomerase inhbitors. Topo I removes DNA supercoils by cleaving a single DNA strand. Torsional stress drives the uncoiling about the intact DNA strand. After the removal of a random number of supercoils, a ligation reaction restores the DNA backbone. Camptothecin intercalates into the nick generated by Topo I and significantly hinders topoisomerase-mediated DNA relaxation [153, 154]. Consequently, in the presence of CPT, the negative supercoiling should increase at the upstream promoter regions bound by the enzyme. If the relationship between transcription and supercoiling is as hypothesized, then the CLDs of the upstream regions from medium expressed genes that depend on Topo I should be more sensitive to CPT

than the CLDs of the highly expressed genes that recruit Topo II. Indeed, treatment of cells with the drug for 5 minutes scaled-up the CLD at the TSS and throughout the upstream region indicating that Topo I activity is generally recruited at promoter regions to control dynamic supercoiling (Fig. 5.5a). The effect of CPT was especially strong for low and medium expressed genes in comparison with high expressed genes (Fig. 5.5b,c and data not shown). The short time of drug administration insures that the CLD profile is the reflection of changes in DNA topology and it is not result of secondary effects [155].

To observe the composite influence of topoisomerases in solving topological problems of transcription, $\beta$-LAP was used. This drug inhibits both Topo I and Topo II, although via different mechanisms. The interaction of $\beta$-LAP with Topo I inhibits the reaction cycle prior to strand cleavage leaving both strands intact [150]. In contrast, poisoning of Topo II by $\beta$-LAP results in the accumulation of covalent DNA-topoisomerase complexes [151]. The precise mechanism by which enzyme inhibitor accomplishes this action is not well understood [156]. The current model postulates that in order to produce a double-stranded DNA breaks [157, 69]. Thus, at low drug concentration (short time treatment) each individual $\beta$-LAP molecule is stabilizing a strand-specific nick rather than a double-stranded DNA break, and diffusion of torsional stress off these nicks should result in the relaxation of the regions served by the Topo II enzyme. Indeed, after 5 minutes of treatment with $\beta$-LAP, the upstream DNA was uniformly relaxed as evidenced by the minimization of the CLD from the TSS to all upstream points (Fig. 5.6a). Therefore, Topo II action is focused in close proximity to the TSSs and helps to

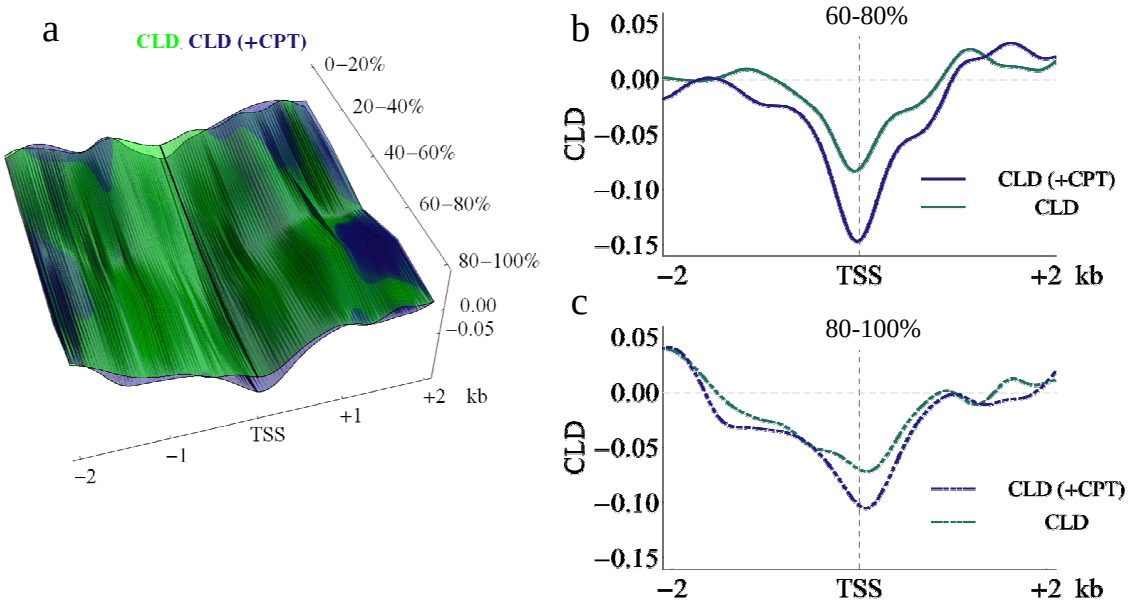


Figure 5.5: Perturbing the distribution of supercoils with camptothecin reveals the pattern of Topo I recruitment to TSSs. (a) 3-D representation of the CLD profiles of genes ranked according to their level of expression in the absence of inhibitors (green surface) and after treatment of cells with CPT (blue surface). (b) Comparison of CLD curves of 60-80% (b) and 80-100% (c)expressed genes in a 4 kb region around the TSS in the absence or presence of CPT. CLD(+CPT) = CL(+DRB) CL(+CPT).

relax negatively supercoiled DNA.

These results show that both topoisomerases are semi-redundantly involved in the relaxation of negative DNA supercoiling upstream of promoters. However, it appears that Topo II is the dominant topoisomerase at the upstream regions of the highly active genes, while Topo I is the dominant topoisomerase at medium output promoters.

## 5.6 Fine tuning of DNA supercoiling with topoisomerase

To conceptualize the role of topoisomerases activity in the steady-state regulation of dynamic supercoils, two scenarios may be hypothesized. In the first, negative torsional stress generated during transcription spreads into the upstream promoter regions (Fig. 5.7a, solid line) where the combined actions of randomly recruited Topo I and Topo II relieve the stress. Because the odds that an upstream region remains topoisomerase-free fall exponentially as distance from the TSS increases, the level of supercoiling should decline in parallel. Alternatively, if topoisomerases are recruited directly to the most dynamically stressed DNA, i.e. TSSs (Fig. 5.7a, dashed line), then level of supercoiling would reduced right at the TSS, but beyond this zone, any residual supercoiling would decay only gradually. The CLD patterns were compared between sets of genes with different expression levels to provide evidence supporting one or the other of these possibilities. Whereas the CLD level in the upstream regions of medium transcribed genes decays exponentially, as expected for the diffuse recruitment of topoisomerases, for high output promoters, the

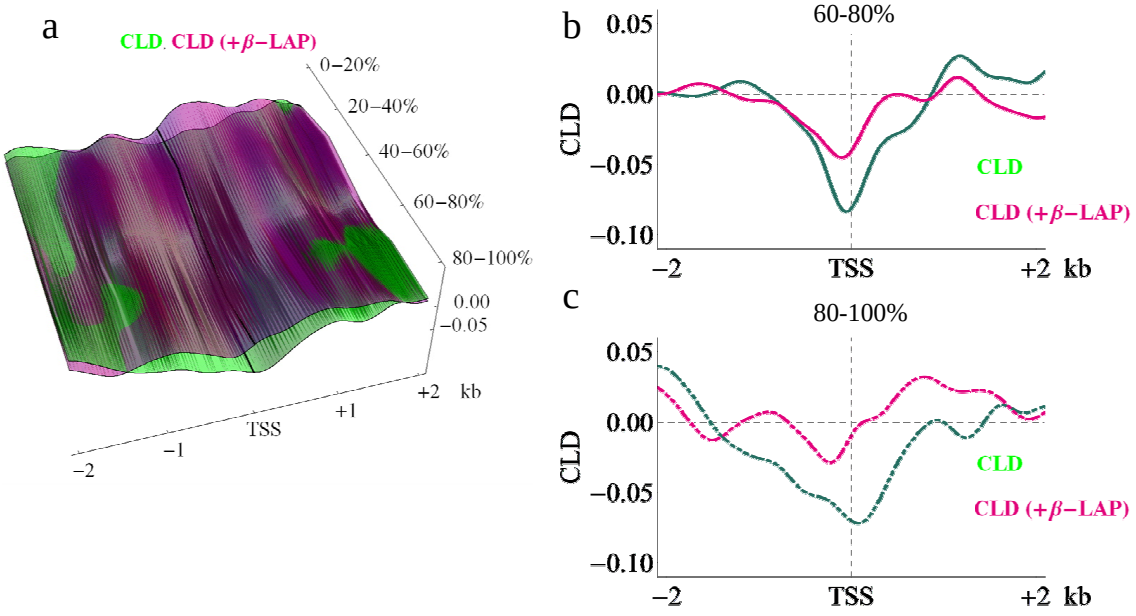


Figure 5.6: Perturbing the distribution of supercoils with $\beta$-lapachone reveals the pattern of Topo II recruitment at TSSs. (a) 3-D representation of CLD profiles over genes ranked according to their level of expression in the absence of inhibitors (green surface) and after treatment of cells with $\beta$-LAP (b - pink surface). Comparison of CLD curves of 60-80% (b) and 80-100% (c) expressed genes in a 4 kb region around the TSS in the absence or presence of drug. CLD(+$\beta$-LAP) = CL(+DRB) CL(+$\beta$-LAP).

CLD declines linearly (Fig. 5.7b). The observed relationship between the rate of transcription, supercoiling intensity, and the response to topoisomerase inhibition suggests that highly active genes require the targeting of a DNA relaxation activity to their TSSs, whereas weakly expressed genes have no such a requirement.

Applying the same logic to experiments using inhibitors selective for one topoisomerase or the other, would allow an estimation of the relative contributions of each topoisomerase to DNA relaxation as a function of gene activity. Camptothecin strongly increased the extent of supercoiling within the upstream regions, but preserved the pattern of CLD distribution: at medium expressed genes it was exponential, and at highly expressed genes it was linear (Fig. 5.5). In the presence of $\beta$-lapachone, the linear decay of supercoiling for highly active genes was preserved, but reduced. At medium active genes the Topo II contribution to upstream supercoiling was inferred from the linear decay of the residual CLD (Fig. 5.6). We conclude that topoisomerases I and II act redundantly within the upstream promoter regions of medium expressed genes (Fig. 5.8a), but when transcription increases, the relief of upstream torsional stress is executed by Topo II targeted right to transcription start sites (Fig. 5.8b).

## 5.7 Discussion

The role of dynamic supercoiling in the regulation and execution of genetic transactions has been incompletely described. Although the existence of torsional stress in actively transcribed DNA and chromatin *in vitro* and *in vivo* has been

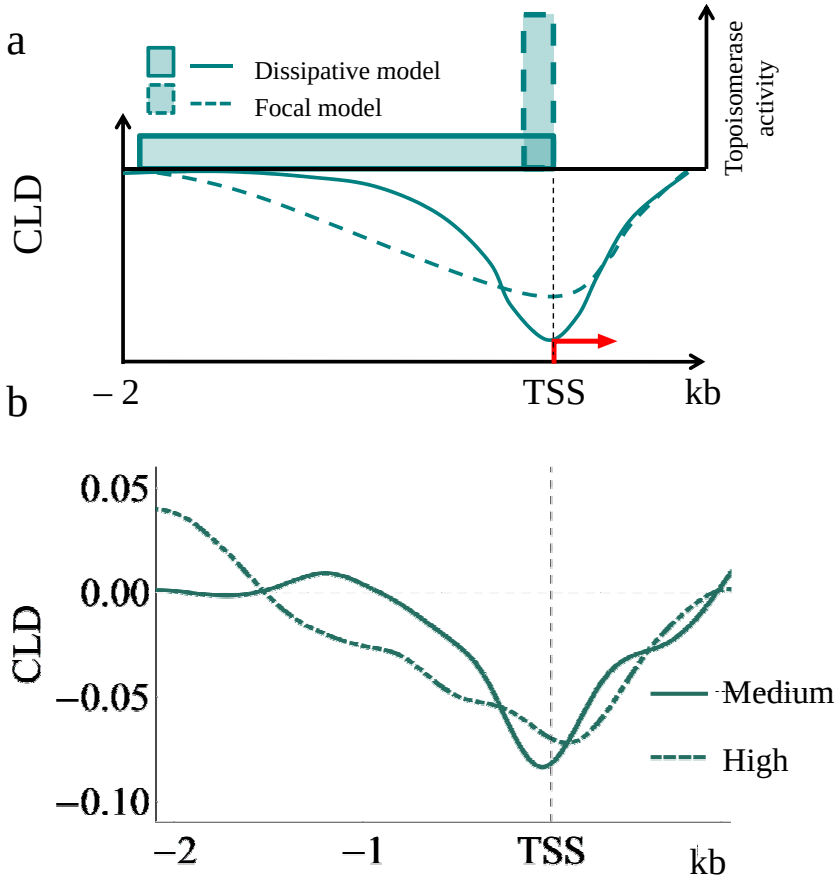


Figure 5.7: Models for topoisomerase recruitment to upstream promoter regions. (a) The focal model (dashed line) hypothesizes that the topoisomerases work close to the TSS and yield a linear decay of superhelical density from the point of topoisomerase binding to DNA. The dissipative model (solid line) postulates that topoisomerases are randomly distributed over the upstream promoter regions, consequently the decay of supercoiling is exponential. (b) Comparison of CLD curves of 60-80% and 80-100% expressed genes in a 4 kb region around TSS.

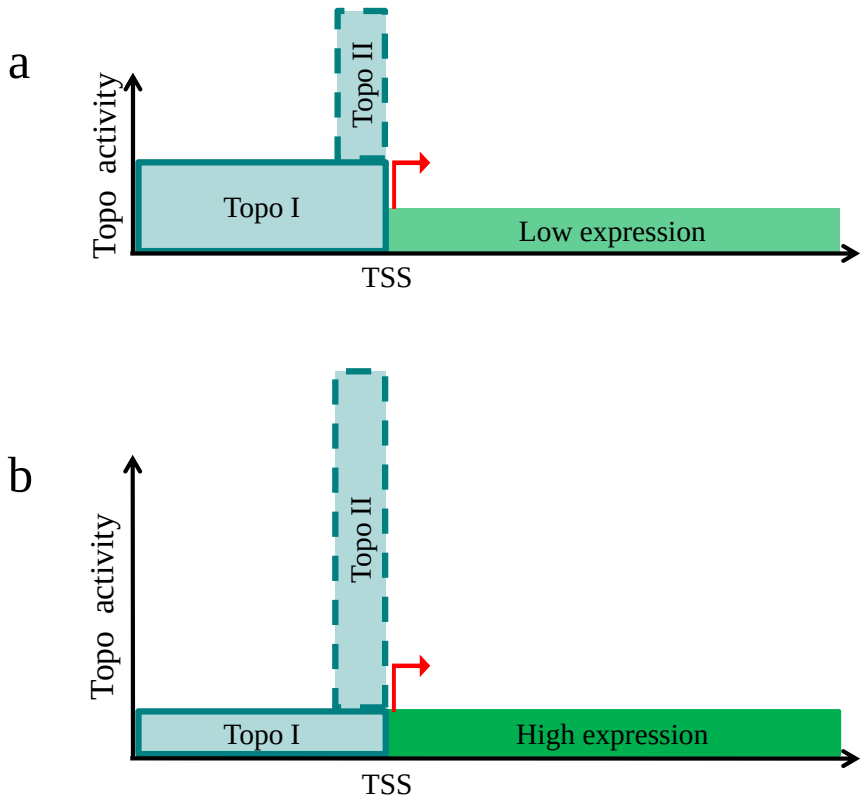


Figure 5.8: Differential topoisomerase I and II utilization in the regulation of transcription-induced torsional stress. (a) From the present results, dynamic supercoiling near low-active genes is managed by topoisomerase I which is distributed over a broad upstream promoter region; (b) whereas highly active promoters recruit topoisomerase II to the focal region near the TSS.

definitively confirmed in systems employing naked or episomal DNAs, respectively, the pervasiveness and significance of dynamic supercoiling for most chromosomal genes has not been established [31, 28]. Recent studies using psoralen as a probe for supercoiling in yeast have revealed variation across large chromosomal territories, but lacked sufficient resolution to relate topology to gene activity because 1) yeast genes and the yeast genome are too compact to confine torsional stress to single targets, and 2) the DNAs immobilized on the microarrays were insufficiently short to enable finer mapping [108]. A genome-wide study of psoralen binding to Drosophila polytene chromosomes was limited by the optical resolution of conventional light microscopy [107]. In the present work, the set of ENCODE genes was studied in their normal chromosomal context, and in the presence of functional topoisomerases. The resolution of high-density oligonucleotide arrays allows us to visualize the fine-grain distribution of dynamic supercoiling near promoters and to reveal its control by topoisomerases. The analysis of these data shows that transcription-generated DNA supercoiling is transmitted locally upstream of promoters, but that highly expressed genes rely upon topoisomerase II to dissipate dynamic supercoiling whereas low-expressed genes depend on topoisomerase I.

### 5.7.1 Modulation of DNA supercoiling

The level of supercoiling depends on two opposing processes: the rapid introduction of torsional stress into DNA, and its removal by topoisomerases or by diffusion into remote regions of the genome [55]. The dynamics of supercoil diffu-

sion should depend on the behavior of chromatin fibers: in principle, the position of individual nucleosomes, the interactions between them, the inter-nucleosomal linker-binding proteins and the nucleosome modifications could all influence supercoil propagation. The data in our analysis reveal that torsional stress is dissipated over a relatively short-range and provide no evidence that dynamic supercoiling butts up against fixed boundaries in chromatin. In such an instance the level of negative supercoiling would be constrained to be a fixed value, decreasing abruptly when crossing the domain border [88]. Alternatively, topological domain boundaries for each particular gene may be heteromorphous or transient in nature, resulting in high variation of domain size between different cells in a population. Although such boundaries would be missed in this analysis, the simplest interpretation of our data is that DNA supercoiling upstream of the active promoter is established mostly by frictional restriction to DNA twist diffusion along the chromatin. Even without fixed boundaries, other architectural features of chromatin could modify the generation and propagation of dynamic supercoils. For example, divergent closely set promoters (which were excluded from this analysis) would drive mutually reinforcing negative DNA supercoils between them. Though the small number of such genes in the ENCODE set does not allow us to investigate in detail the contribution of co-expression on the level of DNA supercoiling in the region between the promoters of divergent genes, the mechanical properties intrinsic to this arrangement suggest that dynamic supercoiling may be an important parameter contributing to co-expression [31].

As suggested by the inhibition of transcription in cells mutant for topoiso-

merases, the relaxation of torsional stress is a prerequisite for efficient transcription [69, 92, 93]. Topo I and Topo II, which can relax both positive and negative supercoiling, are fully redundant for loss of the other in yeast, so only their combined absence severely impairs transcription elongation. However, this is not the case of mammalian cells where Topos can only partially compensate each other suggesting the existence of specific and peculiar functions in the context of transcription. Different topological problems arising during gene activity may dictate specialized roles of each topoisomerase since the positive and negative supercoils generated by transcribing RNA polymerase distort DNA differently and reside in different molecular environments [158]. Accordingly, the differential recruitment of topoisomerases to active genes may be context dependent [141, 94]. The results of this study reveal two characteristics of the relaxation of transcription-induced DNA torsional stress by topoisomerases. First, both Topo I and Topo II prevent the buildup of negative supercoiling in the upstream promoter region. Since Topo I is a torque-sensitive topoisomerase with low activity on the nucleosomal template [96, 97], we infer that it operates mainly downstream of the elongating RNA polymerase where accumulated positive torsional stress and histone modifications make nucleosomes labile [114]. Topo I is well suited for the relief of positive torsional stress that might otherwise stall transcription *in vivo* because it is a processive and "rapid" enzyme that should work well in regions with a high demand for relaxation. Evidently, these regions are downstream of transcription start sites where topoisomerase inhibition can drastically reduce the rate of elongation *in vivo* [75]. Previous results suggest that in yeast Topo II binds to nucleosome-free regions near the transcription

start sites of active genes [95], whereas in mammalian cells binding of the enzyme is enriched near the promoter region of Topo II sensitive genes [159]. In addition, the activity of Topo II would be favored by the crossing of DNA segments [160] as occurs when plectonemes form in DNA with unrestrained negative supercoils that cannot be buffered by chromatin rearrangement [66, 161]. Because all elongating transcription complexes impose a 90-degree bend in the template, as downstream DNA is twisted into the RNA polymerase active site, the upstream DNA exits, it is translationally rotated generating writhe [162, 163]. Therefore, Topo II would be more efficient in relaxing negative supercoiling produced behind of the transcribing RNA polymerase. The twin-supercoiled-domain model predicts that dynamic negative supercoiling is highest at the promoter [48]. Accordingly, the activity of Topo II should be localized near the TSSs of highly active genes as demonstrated in our experiments.

Second, besides draining negative supercoils, it may be important to sustain a steady-state level of torsional stress in upstream regions to manage supercoil-driven structural transitions that serve as a gauge of ongoing transcription [121, 31, 28, 29]. We find that the activity of processive, fast but difficult to control [97], Topo I is reduced at promoters of highly active genes relative to Topo II. DNA relaxation at these genes is accomplished by the step-by-step DNA transport activity of Topo II in which ATP-driven conformational changes of enzyme implement a very transient DNA breaking step [164]. Thus for highly active genes the transient nature of DNA-topoisomerase complexes and topological homeostasis could be enforced by the preferential use of Topo II. Coordinating the rates of transcription and DNA

relaxation adjusts the particular level of DNA supercoiling of many different genes. This conclusion suggests that in contrast to the current paradigm, the relaxing activity is essential not only for solving the severe topological problems arising during transcription but it is required to establish a sturdy level of negative supercoiling within the regulatory regions of active genes.

### 5.7.2 DNA supercoiling in regulatory pathways

In the recent years much evidence has accumulated to support the idea that DNA mechanics serve a variety of regulatory functions [55, 30]. *In vitro* studies suggest that chromatin structure is functionally coupled to DNA topology [114]. DNA supercoiling may assist chromatin remodeling and, by influencing chromatin structure or DNA conformation, may modify DNA-transcription factor interactions [19]. Propagation of torsional stress through DNA may serve as an efficient long-range signal by changing the energy landscape of the chromatin fiber [165]. This signal could restrict or promote the enrollment of DNA conformation-sensitive proteins at regulatory modules [31, 29], or could facilitate protein-DNA interaction over long distances [114]. The same overall regulatory outputs could be achieved only slowly by adjusting the concentrations of particular transcription factors, but DNA supercoiling has the capacity to govern a local specific transaction in real time. Our results reveal that transcription-generated supercoiling has sufficient amplitude and prevalence throughout the genome to modify *cis*-element structure and chromatin conformation to support previously postulated gene regulatory mechanisms.

Finally, we emphasize that because the structure and mechanics of cellular RNA polymerase is conserved across eukaryotes and prokaryotes, then many of the DNA topology-sensitive regulatory mechanisms of transcription in bacteria may also operate in higher organisms. Despite differences in the number and complexity of accessory transcription regulatory components between kingdoms, both of them are forced to contend with the same polymer physics: the requirement to strongly bend DNA for pre-initiation complex formation and the need to locally melt DNA during transcription initiation [163, 166]. Negative supercoiling facilitates both bending and melting [167, 168, 169]; consequently, this fundamental linkage between DNA topology and transcription is maintained in both prokaryotes and eukaryotes [60]. In gyrase-containing bacteria, genomic DNA is globally maintained in an undertwisted state to optimize the transcription of many genes [170], but in higher eukaryotes where genes are often separated by large segments of inactive DNA, each gene may contribute to topological homeostasis at its own promoter. By coordinating the relaxing activity of topoisomerases with the rates of transcription, gene regulatory regions are kept at the constant, but still dynamic, level of DNA supercoiling. As a consequence of this functional supercoiling the early rate-limiting steps in the transcription process might be modified to allow more efficient production of RNA [166]. The complexity of transcriptional regulatory processes in eukaryotes in comparison with bacteria and the short range of torsional stress propagation insure independent topological regulation of different genes. As costs decline, this psoralen-based procedure for the analysis of DNA topology may adapted for NextGen sequencing and may help to uncover other DNA topology-related mechanisms in genome

functioning.

## 5.8 Methods

### 5.8.1 Cell culture

Raji cells were synchronized in early G1 phase of the cell cycle by treatment with 1.5% (v/v) DMSO for 96 hours. Cells were released from DMSO in fresh medium and experiments were conducted 6 hours later. When indicated, the gene transcription was inhibited using $40\,\mu M$ of DRB (or 5,6-dichloro-1-$\beta$-D-ribofuranosylbenzimidazole) for 30 minutes. To inhibit topoisomerases, cells were exposed to $10\,\mu M$ $\beta$-lapachone or camptothecin for 5 min.

### 5.8.2 Psoralen photobinding assay

$2 \times 10^7$ cells per 10 ml of media were treated with 140 $\mu$L of a saturated solution of 4,5',8-trimethylpsoralen in ethanol for 4 min at 37°C. To photocross-link the DNA strands, plates with cells were exposed to $3.6 kJ/m^2$ of 365 nm light (ultraviolet lamp, model B-100 A, Ultra-Violet Products). crosslinked genomic DNA was isolated by RNase and Proteinase K treatment in lysis buffer (10 mM Tris-Cl pH 8.0, 100 mM EDTA, 0.5% SDS), followed by repetitive phenol/chloroform extraction and ethanol precipitation. Purified DNA was sonicated (Sonicator, Ultrasonic processor XL, MISONIX Inc. at 15% of power) to produce 250 bp average-size DNA fragments. DNA was then heat-denaturated and incubated at 55°C for 1 hour in glyoxal buffer. Glyoxylated non-crosslinked and crosslinked DNA fragments were

separated by electrophoresis (3% agarose gel electrophoresis in 10 mM sodium phosphate buffer (pH 7.0) at 2 volt/cm for 12 hours). With this protocol, the ratio of non-crosslinked DNA to crosslinked fragments is 3 to 1. After electrophoresis, the gel was incubated with denaturing solution (0.5 M NaOH, 1.5 M NaCl) at 65°C for 3 hours to reverse psoralen crosslinks and stained with SYBR-green [107]. Crosslinked and non-crosslinked DNA fragments were purified by electroelution and hybridized in different combination to Nimblegen ENCODE arrays (50-mer probes tiled with 12-bp overlap across non-RepeatMasked regions of ENCODE, plus $100\,kb$ region around c-myc gene). Three biological replicates with hybridization to new array for each were conducted for all experiments. DNA labeling, hybridization, detection, data extraction and quality assessment were performed at NimbleGen.

### 5.8.3 Gene expression assay

Nuclear RNA was prepared using the Qiagen RNeasy kit. RNA was isolated from the nuclear pellets resuspended in the kit lysis buffer and processed according to the protocol. RNA was converted into double-stranded DNA by using SuperScript Choice System for cDNA synthesis (Invitrogen). cDNAs were sonicated to average fragments of 250 base pairs and hybridized to Nimblegene ENCODE arrays together with genomic DNA sonicated to similar size. In total, three biological replicates with a new array for each were performed. Data were generated at NimbleGen. Expression levels were defined as the average signal at the annotated gene, normalized by the number of probes (see section 6.4 for more details).

### 5.8.4 Chromatin Immunoprecipitation (ChIP) for microarray

ChIP assays were performed with Raji cells as described with minor changes [171]. Briefly, $5 \times 10^7$ cells were crosslinked with 1% formaldehyde and sonicated in TE to produce chromatin fragments of 800 bp on average. Immunoprecipitations were carried out using 4 $\mu$g of antibodies. For qPCR detection, the percent of IP enrichment as compared to input was calculated using FastStart DNA Master SYBR Green I kit (Roche Diagnostics) and data are presented as the fold change with respect to a negative region of drug treated cells. Nine genes were analyzed in total; three genes in each group ranked according to the RNA production: 0-5%; 55-60%; 95-100%. All detection primers are listed in table 5.1 and complete protocol are presented in section 5.8.5.

### 5.8.5 Chromatin Immunoprecipitation (ChIP) & QPCR for Topo treatments

Chromatin Immunoprecipitation (ChIP) samples were prepared from Raji cells following Barsky et al. protocol with minor changes [171]. Briefly, $5 \times 10^7$ cells were cross-linked with 1% formaldehyde for $10\,min$ at $37\,^{\circ}$C. Cross-linking was stopped by the addition of glycine to $125\,mM$ final concentration and cells were washed twice with PBS. After harvesting cells by scraping, the pellet was washed once with PBS plus 0.5% BSA and resuspended in TE ($10\,mM$ Tris-HCl $pH\,8.0, 1\,mM$ EDTA $pH\,8.0$) to a final concentration of $1 \times 10^6\,cells/ml$. Samples were sonicated 20 times with $20\,sec$ pulses, $30\,sec$ resting, using the Ultrasonic Processor XL (HEAT

Table 5.1: List of all detection primers used for ChIP and QPCR

| Name | Fw Primer 5'>3' | Rv Primer 5'>3' |
|---|---|---|
| *CKMT1A* | GCATTCATTCTCCTTGCTACC | GAGAGTAAAGGCGAGTGGTGTA |
| *TUFT1* | TAAGGCAATGTGTCCCGC | GAAAGGCAGGCACCAAGG |
| *CTAG2* | CTGGGTTCGGCAGTATCAGT | CCTTTCCTGTGGATCTGACC |
| *PFTK1* | CAAAATAAGGCACCCTACATCTG | GAGTCCAGTTGTTTGAGCGG |
| *HISPPD2A* | CTTGATGCTCCCTTCCTTTG | GCACAAACTCTGCCTCTTCC |
| *MIER3* | AGGAATGGGAGATGGAGACC | TTCTCTGCCCTGTCGATCTT |
| *MYC* | GGACTCAGTCTGGGTGGAAGG | AAGGAGGAAAACGATGCCTAGA |
| *IRF1* | GGGAGGGTTTCAGTCCTAGC | CCATCACAGCAAACCATCAA |
| *UQCRQ* | TGTGGCTGAAACTGACGAAAC | AGCACCAAAATCAGGGACAC |
| *NT* | GCAGTTCAACCTACAAGCCAATAGAC | CACAAATTAGCGCATTGCCTGA |

System) to produce chromatin fragments of 800 $bp$ on average. After clarification by centrifugation, sonicated extracts were adjusted to the conditions of RIPA buffer by adding 1% Triton $X100$, 0.1% Na-Deoxycholate, 0.1% SDS and 200 $mM$ NaCl.

4 $\mu$g of anti-Topo I and Topo II were mixed with Dynabeads Protein A (Invitrogen) and incubated at 4 °C for 6 $hr$ with rotation. Chromatin from $5\times10^6$ cells was added to the Protein A-antibody complexes and incubated overnight at 4 °C with rotation. Immunoprecipitates were washed twice with RIPA buffer (10 $mM$ Tris-HCl pH 8.0, 1 $mM$ EDTA $pH$ 8.0, 1% Triton $X100$, 0.1% Na-Deoxycholate, 0.1% SDS, 200 $mM$ NaCl); twice with RIPA buffer plus 300 $mM$ NaCl; twice with LiCl buffer (10 $mM$ Tris-HCl $pH$ 8.0, 1 $mM$ EDTA $pH$ 8.0, 250 $mM$ LiCl, 0.5% NP40, 0.5% Na-Deoxycholate); and twice with TE.

The beads were then resuspended in TE plus 0.25% SDS supplemented with

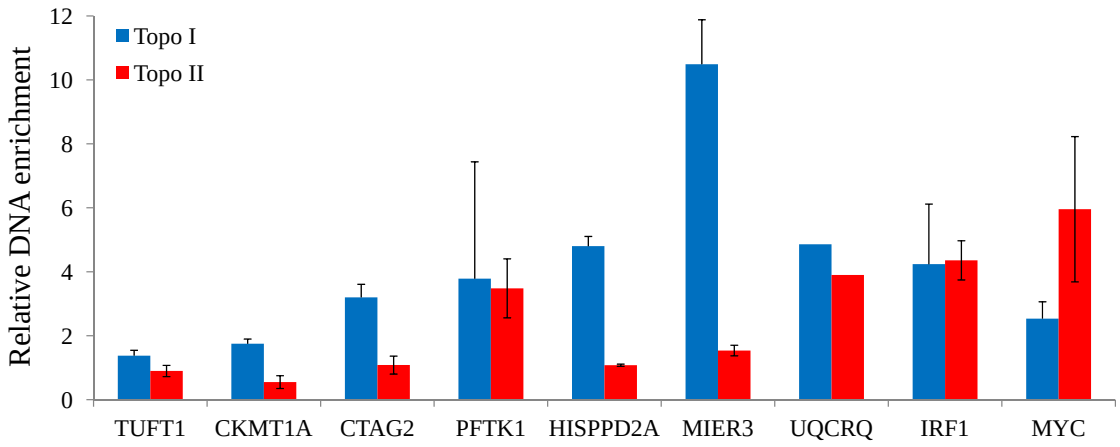


Figure 5.9: Relative enrichment of topoisomerases in promotor regions of genes after treatment with CPT or $\beta$-Lap.

proteinase K (500 $\mu$g/ml, Roche) and incubated overnight at 65 °C. The DNA was recovered from the eluate by phenol chloroform extraction followed by ethanol precipitation in the presence of 4 $\mu$g of glycogen (Roche) and dissolved in TE.

Real-time PCR was performed by using the LightCycler 480 and the SYBR Green I Master kit (Roche Diagnostics). At least four dilutions of genomic DNA were run to generate the standard curve. Quantification and melting curve analyses were performed using the Roche LightCycler software by the crossing point method as indicated by the supplier. For Topo I and Topo II antibodies the DNA recovered values were around 100-fold more enriched than non-immune control. Data in the Fig. 5.9 are presented as relative enrichment of topoisomerases in promotor regions of genes after treatment with CPT or $\beta$-Lap. All detection primers are listed in Table 5.1.

# Chapter 6

# Time series analysis and novel noise prediction methods

This chapter provides details of the mathematical framework developed for signal extraction from our data. I performed all the bioinformatic analysis presented in this chapter.

## 6.1 Overview

In this chapter we'll discuss the various aspects of our microarray data, and the underlying principle / controls. We'll summarize various definitions, analysis methods and test the mathematical model discussed in section 4.6.

First we will discuss some general principals for analysis of time series data that were developed as part of this dissertation. Since our data is noisy, first step is to understand reproducibility of experiments.

## 6.2 Reproducibility

Microarrays have been routinely used for the ChIP-chip experiments, where the enrichment of bound sequences is often 10–100 fold higher than the background. However, for the current series of experiments, namely psoralen intercalation, this is not the case. The maximum observed relative enrichment of psoralen photobinding

under physiological conditions is approximately two fold [103], as the free energy of intercalation of psoralen in negatively supercoiled DNA is much smaller than the corresponding binding energies of typical antibodies. There is also a finite, although smaller, free energy of intercalation in relaxed DNA. Psoralen binding sites are not focal, but are continuously distributed across the genome. As a result the unprocessed data have a very low signal-to-noise ratio (SNR)[1], and conventional methods and standards for mapping molecules bound to DNA are inadequate without modification.

Here we present a method developed to study such low energy / low specificity effects. This method is capable of extracting signal from low SNR data (as low as less than $15^{-2}$), it is unsupervised and has been calibrated.[2] The underlying assumption is that the noise is of much higher frequency than the real signal and it's uncorrelated to the real signal (which in this case is psoralen-binding[3]).

As an example, we define a hypothetical (low frequency) function and overlay increasing levels of white noise[4] (6 replicates). The function was designed so that it has a low frequency signal (based on what we observed from our datasets) and distinctive features of different amplitudes (various peaks and valleys of different amplitudes). For this simulation, the chosen noise levels were in a range that was

[1]See section 6.4.1 for definition.

[2]See section 6.3 for calibration details.

[3]Because we don't expect psoralen intercalation (and level of supercoiling) to change abruptly from one base-pair to next, while the microarray data does show high variation.

[4]Note that although white noise has a flat frequency spectrum (i.e. all frequencies are present) the net frequency component (power) for any given frequency is much smaller than the signal frequency.

much wider than than the observed noise level from the experiment (see section 6.5.1 for more).

The noisy data is then smoothed using Fourier Convolution Smoothing [148], and plotted in Fig. 6.1 along with the raw data, and the original function. We observe that as the noise level increases, the 6 replicates look increasingly different although they are all derived from the same starting function modified by same level of noise. This suggests that when noise levels are high, we cannot ask for reproducibility *from individual experiments*.[5]

To achieve reproducibility/reliability we need to repeat the experiment several times.[6] The number of replicates required depends on the level of noise. If the noise levels are low one or two more experiments suffice. For higher noise levels, higher numbers of replicates are needed.

Let's say that we start with four replicates. These can be subdivided into four subsets of three replicates (by dropping one of them). Now if the average profiles of each subset are similar, then there are enough replicates to make a reliable inference from the data. If the averages are not comparable, that means more replicates are required, and so on.

This is the prescription for a generic case where the actual behavior is not known. For the simulation under discussion we have a direct benchmark for com-

[5]Reproducibility is a fundamental demand of any scientific experiment, and is key for its acceptability and validity. However, under certain stochastic conditions the system can have high degree of variability and exact reproducibility can't be achieved.

[6]Just like one will have to toss a coin several times to test whether it's a fair coin or not, just one or two tosses won't be able to give a definitive answer.

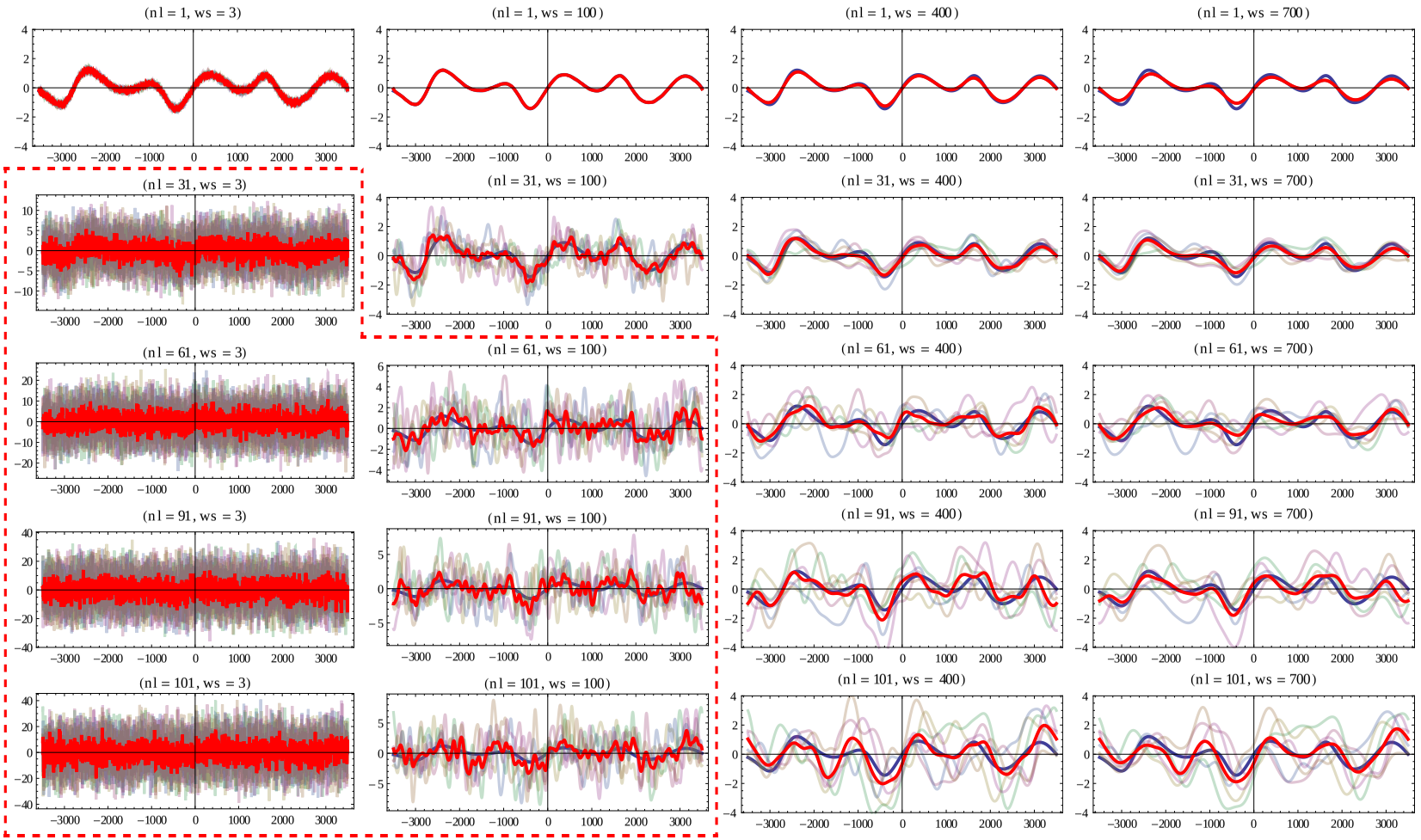


Figure 6.1: Different panels show the same hypothetical response function (in Blue). Each panel has 6 replicates (in dimmed colors) with various noise levels $(nl)$[a] overlayed and smoothed with various window sizes $(ws)$. The average of the replicates is shown in Red. Note that when the noise level is high, the replicates (of the response function with same noise level) behave very differently, but the original behavior is recovered upon averaging. The plots shown in dashed box are on different scales.

[a] See section 6.4.2 for a definition of noise level $(nl)$.

parison, i.e. the original function which was corrupted with different levels of noise. The law of large numbers guarantees an accurate result.

Fig. 6.1 suggests that with the average of 6 replicates, we are able to qualitatively regenerate the original function for SNR as low as $15^{-2}$ (i.e. noise amplitude $\sim$15 times that of the signal amplitude).[7]

If it is not possible to do enormously large number of replicates (due to say economic reasons), the average of all the replicates done is a better measure than the individual experiments. It may seem that a large number of replicates might be needed, but that is not true. For high noise experiments like microarrays, even for our low free energy effect, 3–4 replicates are sufficient to achieve an adequate level of accuracy (with meta-analysis this number comes down to 2–3 experiments).

## 6.3 Calibration for SNR extraction from a given data

The method described in the previous section can be evolved to generate a calibration for estimation of signal-to-noise ratio (SNR) (or noise levels)[8] from a given data provided that the data meets the criterion described in the previous section. To calibrate, we first define a characteristic function based on known features of data. Then we overlay different levels of white noise on this data, which are equivalent to different replicates. At low noise each replicate closely mimics the original function. But as the noise levels go up, the replicates are averaged in different combinations

[7]This is a conservative estimate, as we were able to recover good correlation for up to about 50 times noise with only 6 replicates.

[8]See definition of noise level in section 6.4.2.

of increasing numbers until we get a close fit to the original profile (see Fig. 6.1).

Several thousand simulations were run for various noise levels[9] (ranging between 1 to 100) on unit signal amplitudes[10] with a mix of various small frequencies (which were chosen based on our experimental data). Each of these noisy data set is then smoothed for various window sizes ranging from 400 to 700 (see Table 6.1). The standard deviation of the differences between original noisy dataset and smooth datasets gives a metric for the preselected window sizes. By averaging a large number of entries, coefficient table 6.1 was generated.

Table 6.1: Calibrated correction coefficients for various window sizes.

| Window Size | Coeff |
|---|---|
| 400 | 3.46323 |
| 500 | 3.46300 |
| 600 | 3.46295 |
| 700 | 3.46295 |

This coefficient table is then used to predict the noise levels of any given dataset. This prediction algorithm was tested on several thousands of simulated datasets[11] generated for various noise levels (ranging between 1 to $10^3$) on various signal amplitudes (ranging between $10^{-4}$ to 10) with a mix of various small frequen-

[9] See section 6.5.3 for the protocol used for simulating noisy data.

[10] Signal amplitude is defined as half of the difference between max and min values of all amplitudes.

[11] Each dataset is used alone, no replicates.

cies (which were much larger then experimental data). Table 6.2 summarize the prediction results.

Note that when we have some knowledge about the noise levels, we are able to successfully predict a much broader range, i.e. up to about noise level 40. However, when we have absolutely no knowledge about the noise level, we can still successfully predict the noise levels up to 23. Our meta-analysis data in Fig. 2 and 3 has a noise level of about 13, which is well within the successful prediction range.

The method presented here gives an unsupervised prediction of noise level. A supervised prediction (i.e. with more information about the data) will give better results, but the unsupervised method is sufficient for the present analysis.

This analysis can help predict the number of replicates needed, for a noisy experiment, up to a desired reproducibility-confidence-interval from just one experiment. A simulation on replicates shows that for noise levels at least up to 46, average of three replicates gives high enough noise reduction so that a fourth replicate doesn't add much improvement. This is a reconfirmation that for the purpose of this work 3 replicates are sufficient.

While generalizing this technique, the following facts must be kept in mind. The calibration (and smoothing) is a function of data size and density, frequency spectrum of the data, noise amplitude[12] and frequency etc. Although a complete analytical understanding of the calibration is beyond the scope of this paper, one can safely say that this method will work for very high noise levels for high frequency data also if the sampling frequency is sufficiently high.

[12]The dependence is only on the noise amplitude and not on the signal amplitude.

Table 6.2: Errors in prediction of noise for datasets with known or unknown noise levels.

| Known Noise Level | Stdev ($\sigma$) of Prediction Errors | Unknown Noise Level | Stdev ($\sigma$) of Prediction Errors |
|---|---|---|---|
| 1 | 0 | 1 | 0 |
| 2 | 0 | 2 | 0 |
| 3 | 1 | 3 | 0 |
| 4 | 1 | 4 | 1 |
| 5 | 1 | 5 | 1 |
| 6 | 1 | 6 | 1 |
| 7 | 1 | 7 | 1 |
| 8 | 1 | 8 | 1 |
| 9 | 2 | 9 | 1 |
| 10 | 2 | 10 | 1 |
| 11 | 2 | 11 | 1 |
| 12 | 2 | 12 | 2 |
| 13 | 2 | 13 | 2 |
| 14 | 3 | 14 | 2 |
| 15 | 2 | 15 | 2 |
| 16 | 2 | 16 | 3 |
| 17 | 3 | 17 | 3 |
| 18 | 2 | 18 | 5 |
| 19 | 3 | 19 | 4 |
| 20 | 3 | 20 | 4 |
| 21 | 3 | 21 | 6 |
| 22 | 3 | 22 | 7 |
| 23 | 3 | 23 | 16 |
| 24 | 4 | 24 | 257 |
| 25 | 4 | 25 | 1190 |
| 26 | 3 | | |
| 27 | 4 | | |
| 28 | 4 | | |
| 29 | 6 | | |
| 30 | 5 | | |
| 31 | 5 | | |
| 32 | 6 | | |
| 33 | 6 | | |
| 34 | 10 | | |
| 35 | 8 | | |
| 36 | 8 | | |
| 37 | 9 | | |
| 38 | 7 | | |
| 39 | 10 | | |
| 40 | 12 | | |

## 6.4 General Definitions

### 6.4.1 Signal-to-Noise Ratio

The signal-to-noise ratio is a commonly used term to describe the signal corruption by noise, and is defined as the ratio of signal power to the noise power, see Eq. 6.1, where $A$ is the root mean square amplitude. For more details please see [172].

$$SNR = \frac{P_{signal}}{P_{noise}} = \left(\frac{A_{signal}}{A_{noise}}\right)^2 \tag{6.1}$$

### 6.4.2 Noise Level

The signal-to-noise ratio, as defined in the previous section, has it's origins in electrical engineering where it relates to the ratio of powers in signal and noise. For the convenience of remembering, and ease of intuitive understanding, we define a new term *noise level*. Eq. 6.2 defines the *noise level* in terms of the signal and noise amplitudes ($a$), which are given by the difference between max and min values of the amplitudes.

$$nl = 2\frac{a_{noise}}{a_{signal}} \simeq \frac{2}{\sqrt{SNR}} \tag{6.2}$$

Eq. 6.2 suggests that, a noise level of 10 would mean that the noise amplitude is 5 times larger than the signal amplitude.[13] In other words, one unit of signal is burried in 5 units of noise.

[13] See methods section 6.5.3 for how this definition is used to simulate noisy data.

### 6.4.3 Definition of Sets

| Ratio | Short | Description | Equivalence |
|---|---|---|---|
| $\frac{XL}{nXL}$ | $CL \rightarrow log_2(\frac{XL}{nXL})$ | Relative enrichment of cross−linked DNA (or psoralen intercalation) in untreated (no drug treatment) Raji B cells | Psoralen binding due to a combined effects of sequence, inherent chromatin structure and transcriptionally generated dynamic supercoiling |
| $\frac{XL\,(DRB)}{nXL\,(DRB)}$ | $CL(DRB) \rightarrow log_2(\frac{XL\,(DRB)}{nXL\,(DRB)})$ | Relative enrichment of cross−linked DNA (or psoralen interercalation) in DRB treated cells | Psoralen binding mainly due to sequence and inherent chromatin structure (DRB would inhibit transcription, so no dynamic supercoiling) |
| | $CLD \rightarrow CL(DRB) - CL$ | | Transcription generated dynamic DNA supercoiling (due to ongoing transcription) |
| $\frac{XL\,(CPT)}{nXL\,(CPT)}$ | $CL(CPT) \rightarrow log_2(\frac{XL\,(CPT)}{nXL\,(CPT)})$ | Relative enrichment of cross−linked DNA (or psoralen intercalation) in camptothecin treated cells | |
| | $CLD(CPT) \rightarrow$ $CL(DRB) - CL(CPT)$ | | Transcription generated dynamic DNA supercoiling in cells treated with CPT |
| $\frac{XL\,(\beta\ Lap)}{nXL\,(\beta\ Lap)}$ | $CL(\beta\ Lap) \rightarrow log_2(\frac{XL\,(\beta\ Lap)}{nXL\,(\beta\ Lap)})$ | Relative enrichment of cross−linked DNA (or psoralen intercalation) in $\beta$−lapachone treated cells | |
| | $CLD(\beta\ Lap) \rightarrow$ $CL(DRB) - CL(\beta\ Lap)$ | | Transcription generated dynamic DNA supercoiling in cells treated with $\beta$ Lap |

### 6.4.4 Meta Analysis

During meta-analysis we average multiple transcribed regions by aligning transcribed regions at the transcription start sites ($\pm 8000\,bp$). For all our analysis, we have averaged the raw data, and smoothed only the final average. The ratios are calculated for each individual probe of microarray.

### 6.4.5 Expression Levels

Expression levels were defined as the average of the scores (or signal) for all probes of an annotated gene body.[14] We had 3 replicates of the expression array hybridizations, and average of expression levels from these three experiments were used for further calculations. The expression level is calculated from raw data which was baseline shifted (no smoothing).

### 6.4.6 Expression Level Classes

Once the expression levels were defined, we classified data in several groups (decades, quintiles, quartiles, tertiles etc.). After looking at these different groups, it was apparent that at the level of resolution of our experiments, the data is best viewed in quintiles. For simplicity of explanation, transcribed regions were classified in three categories (based on the expression levels): Low (0–20%, 20–40%), medium (40–60%, 60–80%), high (80–100%).

### 6.4.7 Baseline Shifting

Since we expect ratios to be small,[15] we normalize the entire hybridization experiment so as to bring the overall baseline across the chromosome to zero. This is achieved simply by averaging the ratios of all probes across the chromosomes, and

[14] In other words the total score, normalized by the number of probes.

[15] Because psoralen has a small free energy corresponding to interaction in negatively supercoiled DNA. Moreover, it does have some affinity for intercalation in relaxed DNA as well. Also, see supplementary discussion section 6.2.

subtracting the average from all the probes.

We also used the same concept baseline shifting to remove the sequence dependent bias of psoralen for DNA intercalation.[16]

## 6.5 Analysis Methods

### 6.5.1 Data Analysis

Owing to the small free energy of intercalation of psoralen, the hybrization data was noisy, and had a very small signal to noise ratio.[17] The appearance of the raw data (for all regions) suggested that there was significant high frequency noise (i.e. large variations over short lengths along the DNA). Considering the magnitude of the bending and torsional persistence lengths for DNA $\sim$50–100 $nm$ (about 150–300 $bp$) [19], variation in supercoiling occurring on a much shorter scale is unlikely unless accompanied by a dramatic structural transitions, almost certainly an infrequent phenomenon. Therefore the high frequency fluctuations were attributed to noise.

In order to suppress this noise, we used a technique called Fourier Convolution Smoothing (FCS) to smooth the data. The technique was presented in [148], and is briefly summarized here (for more details please refer [148]).

FCS takes the data set $T$ ordered by abscissa $t$, performs the smoothing on ordinates $d$ and reattaches the abscissa values to the corresponding smoothed ordinate values. Consider an ordered time series data set with $N$ pair of data:

[16] Also see section 6.5.2.

[17] See section 6.2.

$$T = \{(t_i, d_i) : t_i < t_j \ \forall \ i < j \in [1, N]\} \tag{6.3}$$

The ordered set of ordinate values ($d$) in $T$ are given by:

$$D = \{d_i : (t_i, d_i) \in T \ \forall \ i \in [1, N]\} \tag{6.4}$$

Here are some functions that we'll use:

$Ceiling(x)$: greatest integer less than or equal to the number $x$,

$RotateLeft(L)$: cycle elements in list of numbers $L$ $n$-positions to the left,

$Sum(L)$: sum of all the numbers in the list $L$.

We now define another $N$ membered ordered list, $k(S)$, for a positive integer $S$ as:

$$k(S) = \{\exp(-2^{-Sn^2}) : \ \forall \ n \in [Ceiling\left(\frac{N}{2}\right), N - 1 - Ceiling\left(\frac{N}{2}\right)]\} \tag{6.5}$$

Now we define the convolution kernel of index $S$:

$$K(S) = \frac{RotateLeft\left(k(S), Ceiling\left(\frac{N}{2}\right)\right)}{Sum\left(k(S)\right)} \tag{6.6}$$

Now we convolve the ordered ordinate list $D$ with the kernel $K(S)$, using the discrete Fourier transform and inverse transform, to get the new smooth ordered list:

$$D * K(S) = InverseFourier(sqrt(N)\ Fourier(N)\ Fourier(K(S))) \tag{6.7}$$

Combining the ordered list 6.7 with the original ordered abscissae:

$$T(S) = \{(t_i, d^*{}_i) : d^*{}_i \in D * K(S)\ \forall\ i \in [1, N]\} \tag{6.8}$$

$T(S)$ in 6.8 represents the smoothed time series data. The original data can be smoothed for a range of the values of the parameter $S$, and the corresponding smooth data sets $T(S)$ can be obtained. Now the error norm of these ordered data sets are computed w.r.t. moving window average of $T$ with a pre-decided window-size ($ws$), which is the only parameter used for smoothing. The $T(S)$ corresponding to the smallest error is the desired smoothing of $T$.

The benefit of FCS is that it dampens the high frequency noise much more than the low frequency noise. The technique uses moving window average as a reference, as a result of which the local features are not lost during an unsupervised noise reduction.

Our microarrays are designed with each probe having 50 *bp* and a 12 *bp* overlap (i.e. 38 *bp* are unique between successive probes). So for any given region of genome or an individual transcribed region, we have a data density of 38 *bp* per data point (i.e. per probe). While doing the meta-analysis,[18] we align all the transcribed regions on the transcription start sites (TSS). Since the TSS are randomly distributed with respect to probes, for the meta-analysis the data density increases to 1.4 *bp* per data

[18]See section 6.4 for definition.

point. The meta-analysis presented in this study uses a window size of 400 data points (equivalent to 561 *bp*).

Based on the DNA properties, we improvised upon the previously described FCS technique to fit it for our data. The ENCODE data on Nimblegen microarrays was not continuous, so whenever we had a break of 600 *bp* or more (i.e. abt 15 probes), those data points were separated into distinct groups, and smoothed individually. Continuous regions with less than 400 probes were also dropped from individual transcribed regions.

Our Nimblegen ENCODE (*hg*18) microarrays had usable data for a total of 855 transcribed regions. Since many of these regions were overlapping, there was a possibility of over-representing a specific gene. In order to avoid this we identified clusters of transcripts/genes that were overlapping or had a TSS within 50 *bp* of each other; and used only the largest of "transcribed region" from each of these groups. This brought down the total number of transribed regions to 445 (with 415 unique genes). See the list of these transcribed regions in Table A.1.

### 6.5.2 Sequence Dependent Background Correction

These 445 transcribed regions were sorted based on the expression levels[19] and segregated in various quantiles (decades, quintiles, quartiles, tertiles etc.). When meta-analysis[20] was performed for all 445 transcribed regions in these quantiled datasets, we observed a graded difference in baselines for each quantile.[21]

[19]See definition in section 6.4.

[20]See definition in section 6.4.

[21]The low expression quantiles had a higher baseline than the high expressing quantiles.

We wanted to understand this difference, and explain it. It is well known that psoralen has a sequence dependant bias for intercalation in DNA. So we sorted the transcribed regions based on the AT content within $\pm 3000\,bp$ of TSS (instead of sorting them by expression). In the meta-analysis, it was very obvious that the AT-rich transcribed regions had a much higher psoralen intercalation, irrespective of expression level. So we have decided to do an AT content dependent baseline shift for different transcribed regions. To reduce systematic errors, these 445 transcribed regions were divided in 10 groups (each having about 44–45 transcribed regions). Now a correction term, for each of the decades, was calculated by averaging the raw ratios in the flanking regions of $(-8000, -2000)\,bp$ and $(2000, 8000)\,bp$ (about TSS) of the constituent transcribed regions.[22] The data for each of the constituent transcribed regions is then baseline shifted using this correction term to get the corrected data, which is used for further analysis.[23]

### 6.5.3 Addition of Noise Levels in Simulations

There are several ways one could add noise on a pure signal. For our simulations, we used the following protocol for noise addition: For a given dataset and noise level (say $nl$) we generate dataset of equal length such that each point is the

[22] If we had enough data points for all the transcribed regions, we could in principle do a baseline shift based on the flank psoralen profile of each individual gene, but due to lack of continuous data points, we have decided to use the the flanks: $(-8000, -2000)\,bp$ and $(2000, 8000)\,bp$ (about TSS).

[23] All the processing was done on raw data, and smoothing was applied only in final step to remove the high frequency noise.

product of $nl$ and a (pseudo) random number in the range of $-\frac{1}{2}$ and $+\frac{1}{2}$.[24]

## 6.6 Conclusions

This chapter presented the analysis technique that were used for analyzing the data presented in chapter 5. We started with the idea that for highly noisy data reproducibility can be achieved by averaging multiple data sets. Then we presented the calibration of the FCS for our data which allows us to predict the noise level of an unknown time series data set (with noise levels upto 21). For datasets which are known to have low noise level the prediction is good for even higher (i.e. upto about 80). This allowes us to predict the number of desired replicates to achieve statistically significant reproducible results. This is useful as it can potentially save money in repeating costly experiments.

[24] Another possibility could be to use a Gaussian distribution with mean, $\mu = 0$, and standard deviation, $\sigma = nl$.

# Chapter 7

# Novel Normalization And Clustering Analysis of NanoString Data

The results presented in this chapter were part of my collaboration with Dr. Avi Rosenberg (from Dr. Mark Raffeld's lab). The experiments were done either by Avi or post docs from other labs. My contribution is all the data analysis presented in this chapter, including the development of the new normalization and error correction protocol, the clustering analysis to identify similar cancer types, and the significance tables to identify genes that are significantly affected in various treatment and control groups.

## 7.1 Overview

RNAs are the link between the transcription and translation processes. Noise at the transcription level is propagated to translation level by means of RNAs. Transcription level noise can be introduced in several steps during the process of transcription, e.g. chromatin opening, initiation, pausing / stalling / promotor escape, elongation and termination. Apart from these, splicing, pre-processing and stability of mRNA are other critical factors for translational noise.

In order to study these effects, it is important to have a reliable estimate of the number of mRNAs of the entire genome, or at least of a sizable subnetwork.

Such an estimate would be a powerful method for understanding transcriptional / translational noise, as it can potentially define the state of a single cell, or a population of cells. These states can then be compared between diseased and healthy cells / populations.

With developments in scientific research, therapeutic cures (or at least palliative drugs) are becoming available for various cancer molecular subtypes, and we are inching towards personalized medicine [173]. However, for many diseases with genetic disorders, one of the greatest challenges for personalized medicine is a reliable prediction of the specific molecular subtype of the specific disease. If the state of the cell can be defined, like described above, in whole or in part, it would be a great help when sub-classifying patients for a specific type of treatments. While the state defined by complete transcriptome would have much more information, and would be more useful, under current technology it costs a lot of time and money. However, if we are able to identify an appropriate signature set of gene, pathway or subnetwork, that would enable us to classify the molecular subtype of the disease of the specific patients.

For this dissertation, we will present some results from our pilot experiments. Instead of counting all the mRNA transcripts from a single cell / cell population / tumor, we can get an estimate of mRNA copies of a set of pre-selected transcripts using the NanoString nCounter$^{TM}$ assay system [36].

Based on our test of the NanoString system, using a subnetwork of 519 human kinases[1] (out of about $>$ 2000 total kinases), we are able to successfully predict the

[1]Kinases are enzymes that transfer phosphate groups from high-energy donor molecules, such

molecular subtypes in a set of 22 lung cancel cell lines (control and treatments). We are also able to identify a large number of significant contributor genes deregulated due to the disorder. The most significant contributors have been widely reported in previous studies in the literature, however we have identified a large number of novel gene targets that have not been reported previously. We are currently testing these predictions.

## 7.2 Introduction

Kinases are present in numerous critical, activated pathways in cancer in general. In particular, for lung cancer, multiple such pathways have been identified. These include $EGFR$, $PDGFR$ and $ALK$ [174]. Changes in these pathways have been proven to be critical in patient care as therapeutics targeting specific activities have become available (e.g. *imatinib* [175], *crizotinib* [176]). Therefore comprehensive understanding of these targets and their expression patterns may prove critical in the next generation of clinical trials of these potent drugs.

In this preliminary report we present analysis of 22 lung cancer cell lines (control groups and treatment groups) analyzed on Human Kinase codeset from NanoString nCounter assay system. The normalized data is analyzed in various ways to get significant and useful insights about clustering of various molecular subtypes of lung cancer, and the functional information about various drugs and their target genes. Here is a brief description of the technology and the methods.

as ATP, to specific substrates, a process referred to as phosphorylation.

### 7.2.1 Summary of NanoString assay

NanoString assay is a very sensitive technique, which can detect RNA transcripts at concentrations as low as 1 copy per 5 cells [177].[2] NanoString can measure as little as 1.2-fold changes of a single transcript at 20 copies per cell ($10\,fM$) with statistical significance ($p < 0.05$) [178].[3]

Such high sensitivity is a great asset for detection of low copy number transcripts as well as small variation in transcript counts. However, due to high sensitivity, the reproducibility of experiments becomes a problem as the experiment is very sensitive to small variation in sample preparations and handling, as well as to non-specific binding of some probes. Thus, small variations in any of these parameters may cause large distortions in the data.

> One of the significant advantage of NanoString assay is that an amplification step is not required. In traditional RNA library preparation techniques, an amplification step is present [179] which introduces sequence specific bias, and hence the results are skewed.

The exact details of the assay are proprietary, but here is a brief summary [177, 180, 181]. See Fig. 7.1. The assay has a predefined set of molecular barcodes that enable the detection and counting of mRNA molecules in complex biological samples. Spatial permutations of florescent probes create a huge diversity of color codes, each attached to a single target specific probe, so each different code represents a different

[2] Based on an input of $100\,ng$ of total RNA.

[3] For genes expressed at levels between 0.5 and 20 copies per cell, detction of 1.5-fold differences in expression levels with the same level of confidence is reported [178].

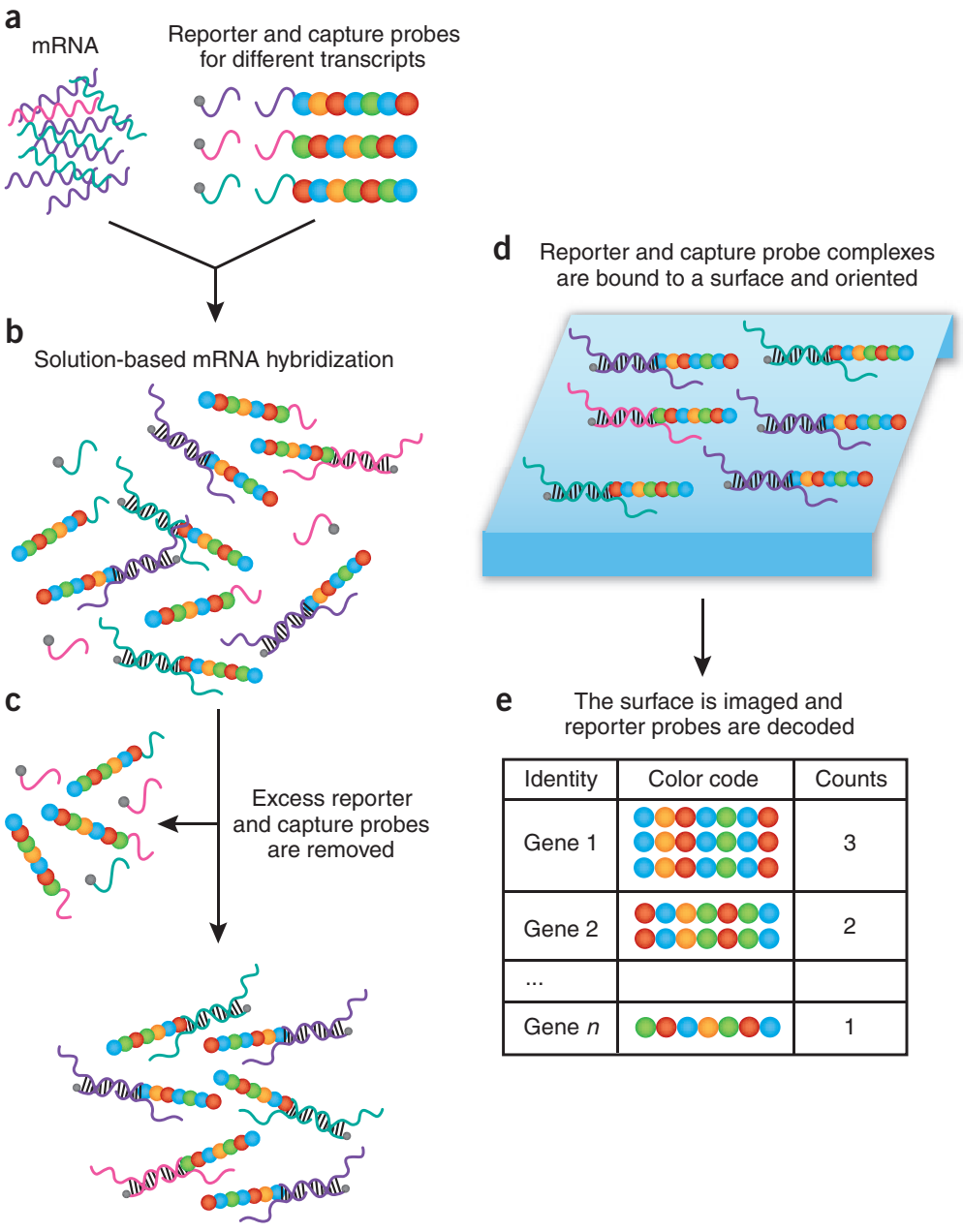


Figure 7.1: Overview of the digital mRNA profiling technology. (a) Total RNA is mixed directly with nCounter reporter and capture probes. No cDNA synthesis or amplification of the target is required. (b–d) After hybridization (b), excess reporters and capture probes are removed (c) and the purified ternary complexes are bound to the imaging surface, elongated and immobilized (d). (e) Reporter probes, representing individual copies of mRNA, are tabulated for each gene. For our experiment, 519 different genes are multiplexed in a single reaction. (Reprinted by permission from Macmillan Publishers Ltd: Nature Biotechnology [180], copyright 2008.)

type of target molecule. Because these probes will bind to the target mRNAs in a one to one ratio, the corresponding barcodes are sorted by their respective code, individually counted and cross referenced to a target identity, yielding a digital count of the target molecules present in the given sample. There are additional gene specific capture probes that work together with reporter probes (attached to barcodes) to increase sensitivity and specificity. After washing away the excess probes, purified probes are loaded into a sample cartridge where they randomly bind to a surface. Current passing through the conducting surface aligns them to the surface. Using a microscope and CCD camera, the data is collected for hundreds of thousands of bound barcodes. (A more detailed description can be found here [177, 180, 181].)

### 7.2.2 Cell line subgroups

Having cancer is like having a car accident. Each accident is different, and could result in injuries of different kinds. Although, for communication purposes, these injuries are usually classified in broad subgroups, e.g. head injury, leg injuries etc., each of these broad categories can mean a number of things depending on the details of the specific injury or patient. Same is true for cancer. When someone develops a tumor (i.e. an "injury") in some part of the body, say lung, this can happen due to mutations (i.e. "accidents") in one or more underlying pathways [182, 183].

Our studies mainly focused on tumors of the lung. We are dealing with a few

specific molecular subtypes:

$EGFR$ mutants: $EGFR$ (epidermal growth factor receptor) is a cell surface receptor of the extracellular protein ligands[4] [184]. Mutations resulting in upregulation (i.e. overexpression) of $EGFR$ has been implicated in various types of cancers, including lung cancer [185]. Based on several studies using multiple drugs, it has been found that patients positive for $EGFR$ mutation have a staggering 60% response rate to the known treatments (which is very high when compared to the conventional chemotherapy alone) [186].

$Kras$ mutants: $Kras$ is a GTPase (i.e. cleaves of the terminal phosphastes from GTP) which also acts as an on/off switch for recruitment and activation of growth factors and receptor signals. Mutation resulting from a single nucleotide substitution causes permanent activation, resulting in various malignancies [185]. It has been found that $Kras$ mutations are present in $EGFR$ negative patients [187]. Although trials are running, but presently there are no drugs effective in treatment of $Kras$ tumors.

$EML4-ALK$ fusion mutants: The fusion of $EML4$ and $ALK$ genes, results in a fusion protein $EML4-ALK$, which was recently identified and implicated in lung cancer [188]. Screening of $EML4-ALK$ has not been standardized yet [189].

$Cripto-1$ mutatnts: $Cripto-1$ is an $EGF$ related gene that codes for a peptide

[4]Ligands are messenger molecules that trigger specific signals by binding to specific sites in the receptor molecules.

growth factor [190]. In human genome, the gene is situated very close to a region that is routinely deleted in a variety of cancers, including lung cancer [191].

Here is a list of the various lung cancer molecular subtypes used in our experiment:

1. $Kras$ mutants:
    - $A549$
    - $H358$
    - $H2122$
2. $EGFR$ mutants:
    - $H3255$
    - $H827$
    - $H1975$
3. $EML4 - ALK$:
    - $H2228$
    - $BEAS2B$
4. $EGFR/Kras$ wild-type:
    - $H322$
    - $H1703$

5. $EGFR$ mutant $\pm Cripto$:

    - $H827Cripto$

Some of these molecular subtypes of lung cancers have been known for sometime, and already have known drugs targetting them. The list below summarizes the drugs used in our study:

*Crizotinib*: *Crizotinib* (or *Criz*) is an inhibitor of $ALK$ and functions by "competitive binding within the ATP-binding pocket of target kinases" [192]. It is currently undergoing clinical trials for lung cancer in adults and children.

$NMS$: $NMS$ is a novel drug that inhibits kinase $PLK1$, which causes apoptosis in cancer cell lines via a potent mitotic cell-cycle arrest [193].

*Erlotinib*: *Erlotinib* reversibly inhibits tyrosine kinases which are highly expressed and often mutated in various types of $EGFR$ positive cancers [194]. So in other words it's an $EGFR$ inhibitor.

We present results for a total of 22 experiments here. Although our experiments didn't have any identical replicates, there were several cell lines that could be grouped by mutation status or by treatments. Below is a list of various control/treatment groups:

1. $EGFR/Kras$ wild-type treatment group:

    - $H3122DMSO$ - control
    - $H3122Criz$ - *crizotinib* treatment

- $H3211NMS$ - $NMS$ treatment

2. $EML4-ALK$ wild-type treatment group:

   - $H2228DMSO$ - control
   - $H2228Criz$ - *crizotinib* treatment
   - $H2228NMS$ - $NMS$ treatment

3. $EML4-ALK$ wild-type treatment group:

   - $BEAS2BPar$ - parental (control)
   - $BEAS2BWT$ - $EML4-ALK$ test condition-1 (active signaling)
   - $BEAS2BKR$ - $EML4-ALK$ test condition-2 (kinase dead mutant)

4. $EGFR$ mutant control-treatment group:

   - $H827$ - parental (control)
   - $H827ER20$ - $20\,\mu M$ treatment with drug *erlotinib*
   - $H827ER40$ - $40\,\mu M$ treatment with drug *erlotinib*

## 7.3 Novel normalization and error correction protocol

The highly sensitive NanoString nCounter system is useful for a variety of applications, such as digital counting of miRNA and mRNA transcripts across a large dynamic range, and measuring copy number variation of DNA. However, as stated in earlier, the high sensitivity may cause large distortions in data due to experimental

variables such as small variation in sample preparations and handling, as well as non-specific binding of some probes. This sensitivity skews the underlying results, and making any inferences and predictions becomes very difficult and unreliable.

To overcome this problem, we developed a novel normalization and error correction approach utilizing the built in "stable" house-keeping genes along with the positive and negative controls. In this preliminary report, analysis of NanoString data is presented using the novel protocol to normalize data for a set of 22 lung cancer cell lines (controls and treatments) on Human Kinase codeset. This codeset had markers for 519 (out of a total of $> 2000$) human kinases that are known to be differentially expressed in the human kinome. The data is analyzed in various ways to get significant and useful insights about the clustering of the various molecular subtypes of lung cancer, and the functional information about various targeted drugs and kinase genes that are affected.

To normalize the data, we use the three internal controls that are a part of the experiment for each of the samples, that are as follows:

1. *Positive controls:* Each sample has a set of spike-in data where a set of non-human genes is added to the reaction mix in known quantities of ranging from $0.125\,fM$ to $128\,fM$.

2. *House keeping genes:* There are a total of 8 house keeping (or "stable") genes, whose expressions are not expected to vary much from one cell line to another cell line.

3. *Negative controls:* Lastly there are a bunch of barcodes attached to reporter

probes for genes that do not belong to the mammalian genome. These codes are expected to be absent from the final data.

When several different samples are compared, we see a large variation (over a massive dynamic range) in estimated expression levels of not only the target genes, but positive controls (up to 4-6 fold variation) and house keeping genes (up to 6-8 fold variation). To make any sense out of the data we needed a way to normalize this data so that different samples are on the same scale. Here is a summary of the normalization and error correction steps that we are currently using:

1. *Intra-sample normalization:* We use the spike in data as reference to normalize (i.e. rescale) the data for each sample. This normalization is performed independently on each sample, and brings all the samples to roughly the same scale. A linear fit was very sensitive to variations in the observed levels of each of the positive controls across all ranges. To improve accuracy, a non-linear interpolation was used.

2. *Inter-sample normalization:* All the 8 housekeeping genes had large variation in their estimated quantities. The fold variation was reduced to about 2-3 fold after intra-sample normalization, however the dynamic range still spanned several hundreds of $fM$s. Based on the covariance analysis, we picked up $HPRT1$ and $CLTC$ as the most stable genes. For our analysis, we used $HPRT1$ which has also been reported previously as a stable gene for data normalization purposes [195, 196]. The geometric mean of the $HPRT1$'s expression levels were used for inter-sample normalization [197]. After normalization, all sets have

the same expression level for $HPRT1$.

3. *Error correction:* Lastly, we use the negative controls to define a common least-count (or 'quanta') for the entire set of experiments so that the observed values of the negative control expression levels is zero. We use this least-count as the sensitivity of the instrument and expression levels of all the genes are rounded to this least-count, giving us quantized variation in expression levels.

   The normalized and error corrected data was used for further analysis.

### 7.3.1 Advantages of our scheme

The existing methods use a linear normalization method across the dynamic range of 3 orders of magnitude ($0.125\,fM$ to $128\,fM$). Moreover, the normalization factor is chosen in an *ad-hoc* way, which is dependent on the number of experiments done, and may vary significantly with each new experiment. This is particularly troublesome for the transcripts with small copy number. Lastly, there is no correlation between the data read and the transcripts count.

Our method overcomes each of these drawbacks:

- Each experiment is normalized with respect to it's own set of positive controls, in a non-linear fashion which is sensitive to the specific dynamic range. Thus the normalization is much less sensitive to variation in observed levels of any one value (as in the case of linear fit).

- We use the information about the spike-in concentrations during the normalization. As a result of this built-in calibration, our normalized data has a

direct correlation with the concentration of transcript levels. This is very useful while comparing different experiments, or transcrtipts.

- Since our error correction protocol results in quantized data using the worst case scenario, we can have high confidence in the values. This is an intuitive way to integrate the information about background noise level into the each of the data point, and makes it is easy to compare the levels of same transcripts among several experiments.

## 7.4 Applications and results

NanoString analysis enables us to ask various questions about a large set of tissue types with various drug treatments and stages of growth or development. In this section we analyze two very basic questions.

The very first question for any experiment is that of reproducibility and ability to put together similar experiments. If the litmus test of reproducibility is cleared, one has confidence in the reliability of the data. The next question is to be able to make testable predictions based on the data.

### 7.4.1 Reproducibility and Hierarchical Clustering

If our analysis is correct, and cells activate / utilize similar kinase pathways, we should be able to cluster these mutation statuses and treatment groups together. To answer the question of reproducibility, we use *Hierarchical Clustering*.

There are innumerable ways to cluster the same set of data, each approach

will answer a specific question (or examine the similarity between datasets from a different point of view). For any given set of experiments, we need to identify the correct question. Once the correct question is identified, it can be used again for similar experiments.

The underlying clustering question is based on the distance function between any two pairs of the multidimensional data set, say $(u, v)$. Here[5] are some of the standard ways to calculate the distances:

(a) Euclidean distance, i.e. simply $\text{Norm}(u - v)$,

(b) Squared Euclidean distance, i.e. simply $\text{Norm}(u - v)^2$,

(c) Normalized Euclidean distance, i.e. $\frac{1}{2}\frac{\text{Norm}((u-\text{Mean}(u))-(v-\text{Mean}(v)))^2}{\text{Norm}(u-\text{Mean}(u))^2+\text{Norm}(v-\text{Mean}(v))^2}$,

(d) Manhattan distance, i.e. $\sum |u - v|$,

(e) Chessboard distance, i.e. $\text{Max}|u - v|$,

(f) Correlation distance, i.e. $\frac{(u-\text{Mean}(u)).(v-\text{Mean}(v))}{\text{Norm}(u-\text{Mean}(u))\text{Norm}(v-\text{Mean}(v))}$,

(g) Bray-Curtis distance, i.e. $\frac{\sum |u-v|}{\sum |u+v|}$,

(h) Canberra distance, i.e. $\sum \frac{|u-v|}{|u|+|v|}$,

(i) Cosine distance, i.e. $1 - \frac{u.v}{\text{Norm}(u)\text{Norm}(v)}$.

[5]For this preliminary proof of concept we have used some of the standard distance functions. In future, when we will ask spedific questions about the molecular signature or the specific drug treatment, we'll have to design specific distance functions depending on the underlying mechanism.

In Fig. 7.2 we show the clustering of all the 22 experiments in aforementioned 9 different ways. Each of these is a unique arrangements of the given data, some are more accurate than the others.

If we examine each one of these clusters carefully, the clusterings in Fig. 7.2 (c), (f) and (i) are among the top choices which correspond to distance functions of normalized euclidean distance, correlation distance and directional cosines respectively. Note how nicely various control and treatment groups are clustered together. With more knowledge about these cell lines, we can identify the correct question (a through i), which can then be used for further analysis and future experiments. For each of these questions, a separate distance matrix can be generated to give an easy feel of clustering in the multi-dimensional space. Fig. 7.3 shows clustering for a smaller set of mutant groups (no treatments) for the same set of questions (a-i).

The fact that we are able to cluster together various molecular subtypes suggests that the experiments are reproducible, and can be used to make further inferences.

### 7.4.2 Identifying significantly affected genes in treatment groups

With some confidence about the reproducibility of the experiments, we can use the normalized data for further analysis. In the remaining part of this chapter, we analyze various groups to make predictions of potential kinases as targets for lung cancer therapeutics [174]. In the first part, analysis of controls and drug treatments is presented, along with a list of genes that are significantly affected by

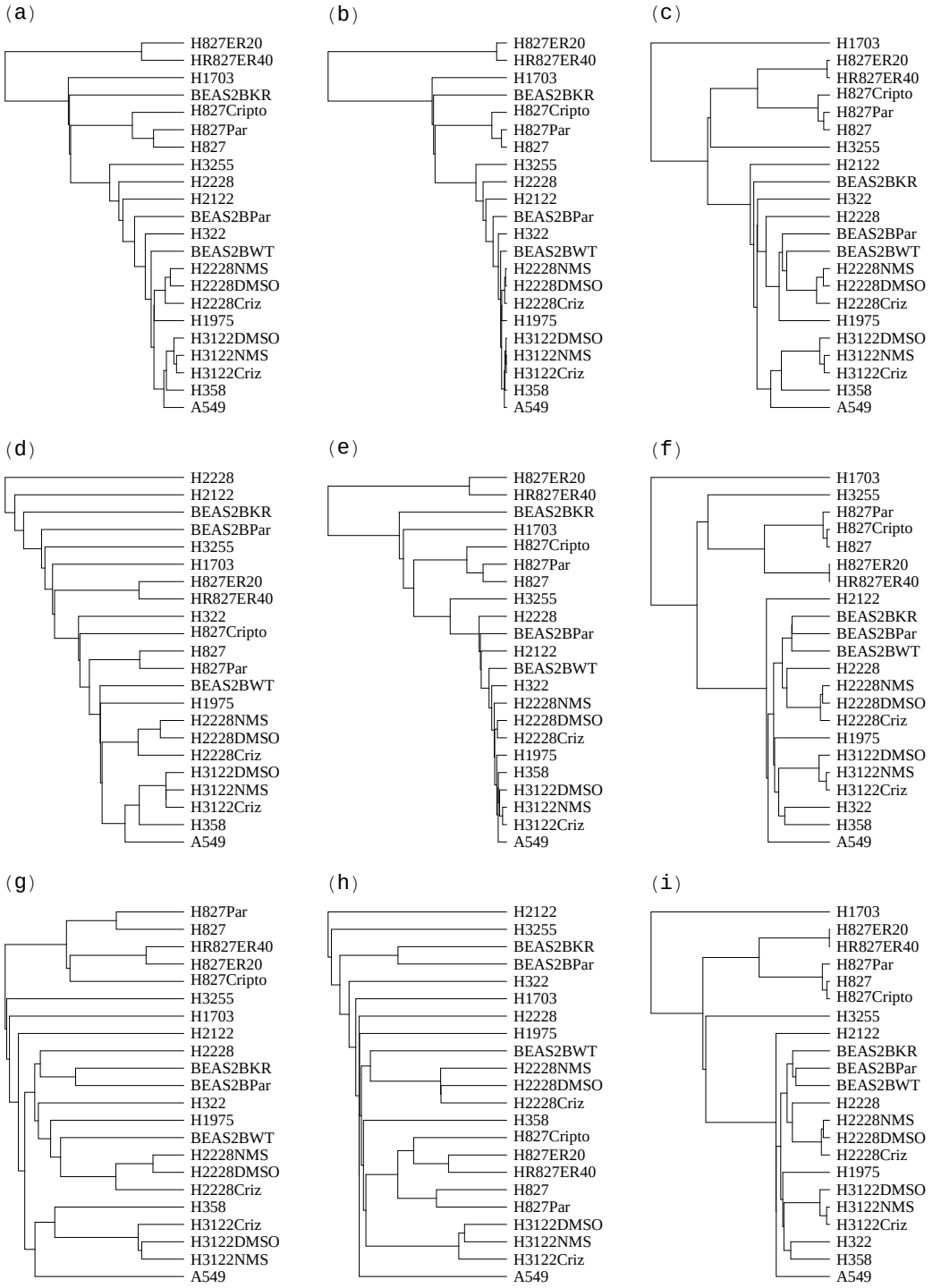


Figure 7.2: Clustering 22 experiments in various ways

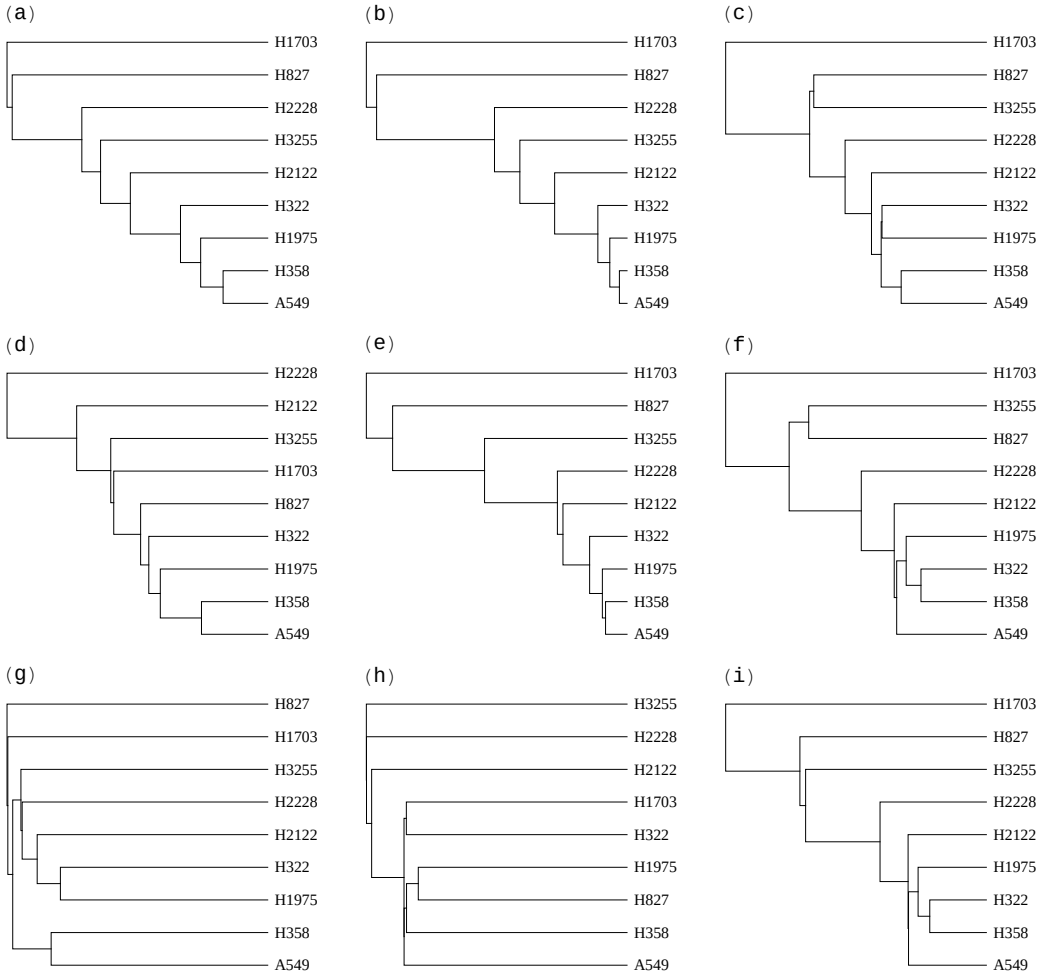


Figure 7.3: Clustering mutant groups in various ways

the treatments. In the second part, various mutant groups are compared to find out the similarities and differences between the affected genes.[6]

We find several interesting features, e.g. we find that *crizotinib* represses most of its target kinases, while $NMS$ activates about half of its targets while repressing the other half (including the common tagets). (See section 7.4.2.1 and Table B.1 & B.2)

#### 7.4.2.1 *Crizotinib* and $NMS$ treatment of $H3122$ cell line

$H3122$ lung cancer cells were treated with DMSO (control) and two drugs, *crizotinib* and $NMS$.

Table B.1 gives the list of genes that are significantly affected by each of the drugs as opposed to the control. This table gives the normalized data, along with pairwise fold changes for the three experiments. The pairs are $1:2$, $1:3$ and $2:3$, and are color coded in red and green depending on whether the fold change is positive or negative. Using the color information, it is very easy to observe that for many of the treatment groups, a significant number of genes show opposite responses to the two drug pairs. See the changing colors in column $1:2$ and $1:3$.

The last column, $Sig$, has significance marker in binary form with 1 indicating significant change in the corresponding pair, while zero representing insignificant

[6]We are only dealing with a limited part of Human Kinome (519 of over 2000), and that we are only studying mRNA expression. There are various other questions which are important (but beyond the scope of this study) such as the stability of mRNAs or proteins, their enzymic activity, effect of various drugs on the wider gene circuits and their feedback to the kinome sub-circuit etc.

change.

> We are choosing a threshold to decide the significanct change. For the current report, significance level was taken to be a minimum difference of 3 quanta or more. This represents really significant changes corresponding to about 5–6 times the standard deviation (i.e. $\sim 5 - 6\sigma$) as comparared to the background levels.

Depending on which of the pairs shows significant change for a given gene, there are following values for $Sig$ (the significance marker): 000, 001, 010, 011, 100, 101, 110, 111. This enumeration simplifies the task of isolating various groups. For example,:[7]

- If we want to look for genes that are significantly affected in both the treatments, we can look for the $Sig$ markers 110 and 111.
- To locate the genes that are affected only by the first drug and not by the second drug, we can look for $Sig$ markers 100 and 101.
- To locate the genes that are affected only by the second drug and not by the first drug, we can look for $Sig$ markers 010 and 011.
- The $Sig$ marker 001 corresponds to a small fraction of genes that are affected

[7]Although this description is true only when there is a comparison between the three experiments, the algorithm can do a pairwise comparison among groups larger than three. Also, this is an analysis where the control is in the first column and treatments are in the remaining two columns.

by the two treatments by small amounts in opposite directions, but the relative change among the treatments is larger and significant.[8]

- The genes corresponding to *Sig* marker 000 are dropped from the list as they do not have significant changes in either of the pairs.

  *Note:* In some cases, where the expression level for a gene is zero in one experiment, while significant in the other experiments, taking ratios gives zeros or infinity. To overcome this problem, the ratios are taken assuming the low expression value to be 100 times smaller than the quanta. This avoids the aforementioned errors, at the same time marking these ratios as significantly different from the other ratios.

In Table B.1, if we look at the genes that are significantly affected by both the drugs (i.e. $Sig = 110/111$), it is interesting to observe that *crizotinib* represses almost all the 58 kinases, while $NMS$ represses about $\frac{2}{3}^{\text{rd}}$ of kinases (see $Sig = 110$ and some in 111) and activates the remaining $\frac{1}{3}^{\text{rd}}$ (see $Sig = 111$), including the common tagets with *crizotinib*.

As stated in the section 7.2.2, *crizotinib* is an inhibitor of $ALK$ and functions by "competitive binding within the ATP-binding pocket of target kinases" [192]. This result suggests that successful inhibition of $ALK$ in the $EGFR/Kras$ mutant cell line $H3122$ causes the repression in expression of about 50 other kinases (Table B.1).

[8] In principle there should not be any genes belonging to this category, but because we are choosing a high threshold for significance, we do see several genes in this category.

Similarly, $NMS$ treatment seems to have a similar repressive effect on almost $\frac{2}{3}^{\text{rd}}$ of these kinases. $NMS$ is a drug that inhibits kinase $PLK1$ [193]. However, if we compare the actual expression levels, it's very evident that the *criz* treatment was much more potent than the $NMS$ treatment.

The results here suggest that either both these drugs have multiple targets, or that their known targets ($ALK$ for *criz*, and $PLK1$ for $NMS$) have functional relations with the affected 58 kinases. (Table B.1)

#### 7.4.2.2 *Crizotinib* and $NMS$ treatment of $H2228$ cell line

$H2228$ lung cancer cells were treated with DMSO (control) and two drugs, *crizotinib* and $NMS$. Table B.2 gives the list of genes that are significantly affected by each of the drugs as opposed to the control. (For details about the table, please see section 7.4.2.1.)

Once again we notice that *crizotinib* represses almost all the kinases it targets, while $NMS$ represses about $\frac{2}{3}^{\text{rd}}$ of kinases (see $Sig = 110$ and some in 111) and activates the remaining $\frac{1}{3}^{\text{rd}}$ (see $Sig = 111$), including the common tagets with *crizotinib*.

As stated in the section 7.2.2, *crizotinib* is an inhibitor of $ALK$ and functions by competitive binding within the ATP-binding pocket of target kinases [192]. This result suggests that successful inhibition of $ALK$ in the $EML4 - ALK$ mutant cell line $H2228$ causes the repression in expression of about 150 other kinases (Table B.2), which is about 3 times larger than the affected number of kinases in

$EGFR/Kras$ wild type cell line $H3122$ (Table B.1).

This means that if $ALK$ is inhibited then the fusion protein $EML4-ALK$ also looses it's functionality. All the kinases that are affected in the $EML4-ALK$ mutant $H2228$ cell line (i.e. Table B.2) but not in the $EGFR/Kras$ cell line $H3122$ (i.e. Table B.1) are targets of the fusion protein.

Similarly, $NMS$ treatment seems to have a similar repressive effect on almost $\frac{2}{3}^{\text{rd}}$ of kinases. $NMS$ is a drug that inhibits kinase $PLK1$ [193]. However, once again, if we compare the actual expression levels, it's very evident that the *criz* treatment was much more potent than the $NMS$ treatment for both $EGFR/Kras$ and $EML4-ALK$.

The results here suggest that either both these drugs have multiple targets, or that their known targets ($ALK$ for *criz*, and $PLK1$ for $NMS$) have functional relations with the affected 58 kinases. (Table B.1)

#### 7.4.2.3 Conditioning on $BEAS2B$ cell line

$EML4-ALK$ mutant $BEAS2B$ lung cancer cells were studied in two conditions: active signaling ($BEAS2BWT$) and kinase dead mutant ($BEAS2BKR$). Table B.3 gives the list of genes that are significantly affected by each of the conditions as opposed to the parental cell line. (For details about the table, please see section 7.4.2.1.)

Note that condition 1 represses almost all observed affected kinases while condition 2 activates almost all of them, including the common targets.

As expected this $EML4-ALK$ mutant cell line $BEAS2B$ shares many deregulated kinases with the previous $EML4-ALK$ mutant cell line $H2228$ (Table B.2). Curiously though, loss of kinase activity (in the untreated kinase dead mutant $BEAS2BKR$ cell line) causes the activation of other kinase pathways, e.g. $MAPK13$, $PAK6$, $ERN1$, $AXL$ and most importantly $EFGR$.

If we compare at the clustering of the untreated kinase dead $EML4-ALK$ mutant cell line $BEAS2BKR$ with the $ALK$ inhibitor treated cell line $H2228Criz$, the two do not cluster together (Fig. 7.2). This suggests that either the same drug is inhibiting multiple kinases, and/or the inhibition of one kinase may have a context dependent effect.

#### 7.4.2.4 *Erlotinib* dosage treatments on $H827$ cell line

$H827$ lung cancer cells were treated by two separate dosages of *erlotinib*: $20\,\mu M$ ($H827ER20$) and $40\,\mu M$ ($H827ER40$). Table B.4 gives the list of genes that are significantly affected by each of the conditions as opposed to the parental cell line. (For details about the table, please see section 7.4.2.1.)

Note that both the concentrations are activating almost all the kinases across the board, although, almost always the smaller dosage is more effective in activation than the larger dosage.

As stated before, *Erlotinib* reversibly inhibits tyrosine kinases which are highly expressed and often mutated in various types of $EGFR$ positive cancers [194]. So in other words it's an $EGFR$ inhibitor.

When the tyrosine kinase is inhibited, we see that $MET$ levels shoot upto about 10-12 folds, and $MET$ is known to have tyrosine kinase activity. So when one tyrosin kinase pathway is shutdown, another route is taken. This assay and analysis tells us the alternate route taken. Note that $MET$ is also an important oncogene which itself is mutated in various cancers.

There are other such examples of kinases, which are upregulated upon *erlotinib* treatment (e.g. $SGK$). Many of these might potentially have some tyrosin kinase activity or a direct relation to one of the tyrosin kinase pathway. The important point to note is that we are able to make a logical guess based on this analysis.

### 7.4.3 Identifying significantly affected genes in mutant groups

In the previous section we identified many genes /pathways that were affected by specific treatments, dosages or experimental conditions. It's no surprise that each of them have a signature that has commonalities and differences from others.

Now, let us focus on the mutant groups alone. The question that we are essentially asking here is how similar (or different) are the different cell lines of the same mutant group? We notice that just as expected from the "accident" analogy, even the members of the same family are different with each other. The net amplitude of the variations is low however.

To draw an analogy from non-linear dynamics, it seems that the internal transcriptome has different "attractor states". A mutation in one of the oncogenes pushes the cell/tumor to a diseased attractor. Many of the molecular subtypes likely

have "small barrier" to switch between these attractors. Thus when one "attractor" is perturbed by means of some treatment, it switches to another "attractor". The significant contributors identified in this section are key kinase contributor which are potentially playing a role in the definition of those attractor states.

#### 7.4.3.1 $Kras$ mutant group comparison

Table B.5 gives the list of genes that are significantly affected by various $Kras$ mutant lung cancer cell lines $A549$, $H358$ and $H2122$. (For details about the table, please see section 7.4.2.1.)

Note that the $H2122$ cell line has the highest activation levels across the board while $A549$ and $H358$ cell lines have a roughly equal distribution of activated and repressed genes between them.

As expected there are many kinases that are expressed significantly in only one of these groups. These kinases, alone or in groups, are potential subclassifiers of the molecular subtypes for each of these specific tumors. (Note that a signature could be the overexpression as well as underexpression.)

The different signature can be picked upfrom the appropriate $Sig$ markers. (For details about the table, please see section 7.4.2.1.)

For $A549$ some contributors are $AXL$, $FGFR$s etc. For $H2122$ some signature contributors are $SNF1LK$, various $STK$s etc.

#### 7.4.3.2 $EGFR$ mutant group comparison

Table B.6 gives the list of genes that are significantly affected by various $EGFR$ mutant lung cancer cell lines $H3255$,$H827$ and $H1975$. (For details about the table, please see section 7.4.2.1.)

Note that for a significant majority of kinases the cell line $H3255$ has the highest level of expression, while the $H827$ cell line has the lowest activation levels across the board.

As expected there are many kinases that are expressed significantly in only one of these groups. These kinases, alone or in groups, are potential subclassifiers of the molecular subtypes for each of these specific tumors. (Note that a signature could be the overexpression as well as underexpression.)

The different signature can be picked upfrom the appropriate $Sig$ markers. (For details about the table, please see section 7.4.2.1.)

#### 7.4.3.3 $EGFR$ and $Kras$ wild type comparison

Table B.7 gives the list of genes that are significantly affected between the $EGFR$ and $Kras$ wild type mutant lung cancer cell lines $H322$ and $H1703$. (For details about the table, please see section 7.4.2.1.)

Note that there is almost an equal mix of genes that are expressed at higher levels in either cell line.

As expected there are many kinases that are expressed significantly in only one of these groups. These kinases, alone or in groups, are potential subclassifiers

of the molecular subtypes for each of these specific tumors. (Note that a signature could be the overexpression as well as underexpression.)

The different signature can be picked upfrom the appropriate *Sig* markers. (For details about the table, please see section 7.4.2.1.)

#### 7.4.3.4 *EGFR* mutant ±*Cripto* comparison

Table B.8 gives the list of genes that are significantly affected by among the *EGFR* mutant *H*827 and its *Cripto* counterpart *H*827*Cripto* cell line. (For details about the table, please see section 7.4.2.1.)

Note that the expression of about 90% of the kinases goes up in *H*827*Cripto* cell line, while the expression of most of the remaining 10% kinases is decreased by about 2-fold (or higher)

As expected there are many kinases that are expressed significantly in only one of these groups. These kinases, alone or in groups, are potential subclassifiers of the molecular subtypes for each of these specific tumors. (Note that a signature could be the overexpression as well as underexpression.)

The different signature can be picked upfrom the appropriate *Sig* markers. (For details about the table, please see section 7.4.2.1.)

### 7.4.4 Summary of predictions

Using the novel normalization and error correction, we have been able to make the not only cluster different moluecular subtypes of lung cancers, but also able to

make significant predictions about the genes that are affected. The table below gives a summary of the predictions.

- From the analysis of $H3122$ and $H2228$ lung cancer cell lines, we find that *crizotinib* causes repression in expression of a majority (50) of the affected genes (58), while $NMS$ activates about half (30) of the affected genes (58) and represses the other half (including their common targets). (See sections 7.4.2.1 and 7.4.2.2 for details.)

- The active signaling (condition 1) represses almost all the affected kinases in $EML4-ALK$ mutant $BEAS2B$ parental cell line, while the kinase dead mutant (condition 2) activates almost all of them, including the common targets. (See section 7.4.2.3 for details.)

- The two chosen *erlotinib* concentrations activate almost all the kinases across the board, although almost always the smaller dosage is more effective in activation than the larger dosage. (See section 7.4.2.4 for details.)

- In the $Kras$ mutant group, the $H2122$ cell line has the highest activation levels across the board, while $A549$ and $H358$ cell lines have a roughly equal distribution of activated and repressed genes between them. (See section 7.4.3.1 for details.)

- In the $EGFR$ mutant group, a significant majority of kinases cell line $H3255$ has the highest level of expression, while the $H827$ cell line has the lowest activation levels across the board. (See sections 7.4.3.2 and 7.4.3.3 for details.)

- In the $EGFR$ mutant cell line $H827$, the expression of about 90% of the kinases goes up in presence of $Cripto$, i.e. $H827Cripto$ cell line, while the expression of most of the remaining 10% of the kinases is decreased by about 2-fold (or higher). (See section 7.4.3.4 for details.)

## 7.5 Future directions

The high-sensitivity of the NanoString nCounter system enables us to make very fine measurements of mRNA expression levels of various genes. Using the new normalization protocol (section 7.3), we make a host of significant predictions that are further testable by means of standard techniques.

Apart from successfully predicting the molecular subtypes of various lung cancer cell lines, there is a wealth of functional information in each of the significance comparison tables that we hope to harness in future work.

Here is a summary of the immediate next steps:

1. Comparison to other techniques of data normalization [36].

2. Choose the correct clustering question that gives us the correct mutation status (see section 7.4.1 for more details).

3. Do the clustering analysis for experiments with the most significant genes dropped.

4. Validation of the predicted, repressed and activated, kinase targets by PCR and other standard techniques.

5. Application of normalization strategy (section 7.3) to more complex samples, e.g. tumors where several different tissue types are mixed.

6. To infer the functional relations among various kinases.

To draw an analogy from non-linear dynamics, it seems that the internal transcriptome has different "attractor states". A mutation in one of the oncogenes pushes the cell/tumor to a diseased attractor. Many of the molecular subtypes likely have "small barrier" to switch between these attractors. Thus when one "attractor" is perturbed by means of some treatment, it switches to another "attractor". Our long term goal is to be able to predict the correct attractor so as to put the push the cell/tumor from a diseased attractor to the healthy one, or at least to a manageable one.

# Appendix A

# Supplementary material for DNA supercoiling analysis

## A.1 Simulation function

Although the microarray data does show high variation, we don't expect psoralen intercalation (and level of supercoiling) to change abruptly from one base-pair to next.

The function used in Fig. 6.1 was designed so that it has a low frequency signal (based on what we observed from our datasets) and distinctive features of different amplitudes (various peaks and valleys of different amplitudes). The underlying assumption is that the noise is of much higher frequency than the real signal and it's uncorrelated to the real signal (which is this case of psoralen-binding).

Eq. A.1 shows the function used for simulation in Fig. 6.1 (range: $\pm 3500$):

$$\frac{\mathrm{Sin}\left(\frac{x}{211}\right)}{2+\mathrm{Sin}\left(\frac{x}{351}\right)}+\frac{\mathrm{Sin}\left(\frac{x}{451}\right)}{2-\mathrm{Cos}\left(\frac{x}{451}\right)} \tag{A.1}$$

It must be emphasized that the scheme will work for any response function with frequencies that are much smaller than the noise frequencies. The smoothing calibration code was tested on several thousands of simulated datasets[1] generated for various noise levels (ranging between 1 to $10^3$) on various signal amplitudes (ranging between $10^{-4}$ to 10) with a mix of various small frequencies.

[1]Each dataset is used alone, no replicates. For more details, see section 6.2 and 6.3.

## A.2 List of genes used

Here is the list of these transcribed regions used for analysis:

Table A.1: List of transcribed regions used in analysis

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 1 | 19 | 59107802 | 59137444 | 59284 |
| 2 | 19 | 59107882 | 59138080 | 59284 |
| 3 | 20 | 33573306 | 33574658 | 343705 |
| 4 | 6 | 74135002 | 74136236 | 441161 |
| 5 | 20 | 33578212 | 33580862 | 140873 |
| 6 | 6 | 74129120 | 74130615 | 154288 |
| 7 | 6 | 74119507 | 74120674 | 340168 |
| 8 | 15 | 41772375 | 41778712 | 548596 |
| 9 | X | 153152125 | 153176632 | 1527 |
| 10 | X | 153101360 | 153114725 | 2652 |
| 11 | 15 | 41673536 | 41678892 | 548596 |
| 12 | X | 153062933 | 153077705 | 5956 |
| 13 | 20 | 33484240 | 33486662 | 554250 |
| 14 | 20 | 33484562 | 33489441 | 8200 |
| 15 | X | 153533444 | 153535036 | 30848 |
| 16 | 19 | 59927813 | 60070473 | AF285439 |
| 17 | X | 153466704 | 153468263 | 653387 |
| 18 | 1 | 149603404 | 149611805 | 8991 |
| 19 | 20 | 33336947 | 33343639 | 128876 |
| 20 | 20 | 33652037 | 33656662 | 80307 |
| 21 | 1 | 149779404 | 149822683 | 7286 |
| 22 | 20 | 33609920 | 33651008 | 80307 |
| 23 | 11 | 5558682 | 5559690 | 340980 |
| 24 | 6 | 74161191 | 74183791 | 55510 |
| 25 | 6 | 73975313 | 74029640 | 80759 |
| 26 | X | 153556723 | 153632526 | 139716 |
| 27 | X | 153499058 | 153500716 | 246100 |
| 28 | X | 153717262 | 153904192 | 2157 |
| 29 | 11 | 4799191 | 4800220 | 119694 |
| 30 | 19 | 59739981 | 59748862 | 90011 |
| 31 | 11 | 4901179 | 4902145 | 79324 |
| 32 | X | 153177344 | 153211894 | 8277 |
| 33 | X | 153660161 | 153686957 | 4354 |
| 34 | 11 | 4826042 | 4827014 | 119692 |
| 35 | 11 | 5230996 | 5232587 | 3048 |
| 36 | 11 | 5036455 | 5037433 | 119678 |

Continued on next page

**Table A.1 – continued from previous page**

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 37 | 6 | 74191313 | 74218720 | 115004 |
| 38 | 11 | 5109497 | 5110448 | 390054 |
| 39 | 11 | 5329313 | 5330252 | 390058 |
| 40 | 11 | 5024331 | 5025267 | 119679 |
| 41 | 11 | 5400006 | 5400960 | 390061 |
| 42 | 11 | 5177540 | 5178506 | 283111 |
| 43 | 19 | 59265068 | 59269173 | 441864 |
| 44 | 11 | 4923964 | 4924906 | 401666 |
| 45 | 11 | 4932577 | 4933519 | 401667 |
| 46 | 11 | 5522366 | 5523329 | 390067 |
| 47 | 22 | 31080871 | 31087147 | 10738 |
| 48 | 11 | 5492198 | 5494501 | 143630 |
| 49 | 22 | 31085892 | 31097063 | 10737 |
| 50 | 22 | 30875518 | 30885243 | 150297 |
| 51 | 11 | 4746784 | 4747723 | 256892 |
| 52 | 19 | 59187353 | 59207732 | 59285 |
| 53 | 19 | 59077278 | 59102713 | 5582 |
| 54 | 6 | 108593954 | 108616706 | 7101 |
| 55 | 9 | 131123115 | 131127005 | 414318 |
| 56 | 11 | 4859624 | 4860689 | 401665 |
| 57 | 6 | 41411504 | 41426593 | 9436 |
| 58 | 22 | 30769258 | 30836645 | 6523 |
| 59 | 15 | 41597132 | 41611110 | 4130 |
| 60 | 11 | 4965999 | 4970235 | 56547 |
| 61 | 22 | 30916425 | 30930718 | 10739 |
| 62 | 22 | 30944462 | 30981318 | 6527 |
| 63 | X | 152780580 | 152794505 | 3897 |
| 64 | 22 | 31526801 | 31589028 | 7078 |
| 65 | 20 | 33506563 | 33563216 | 11190 |
| 66 | 22 | 31238539 | 31732683 | 8224 |
| 67 | 22 | 31239399 | 31784329 | 8224 |
| 68 | 6 | 41829976 | 41834895 | 647014 |
| 69 | 5 | 131315195 | 131375214 | 23305 |
| 70 | 5 | 131424245 | 131426795 | 3562 |
| 71 | 5 | 131170738 | 131357870 | 23305 |
| 72 | 22 | 31140289 | 31183373 | 254240 |
| 73 | 5 | 131317500 | 131375553 | 23305 |
| 74 | 11 | 4892524 | 4893469 | 81282 |
| 75 | 11 | 4976788 | 4977736 | 119682 |
| 76 | 5 | 141953305 | 142045812 | 2246 |
| 77 | X | 153452672 | 153453380 | 286967 |
| 78 | 22 | 31087313 | 31107216 | 646618 |

Continued on next page

**Table A.1 – continued from previous page**

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 79 | 5 | 131556201 | 131590834 | 8974 |
| 80 | X | 152891434 | 152901834 | 8269 |
| 81 | 11 | 4885175 | 4886114 | 119687 |
| 82 | 11 | 5466512 | 5467469 | 390066 |
| 83 | 22 | 31039083 | 31041792 | 646599 |
| 84 | 11 | 5431294 | 5432233 | 390064 |
| 85 | X | 152853916 | 152863426 | 5973 |
| 86 | 5 | 131905034 | 131907113 | 3567 |
| 87 | 11 | 131033416 | 131038060 | 399980 |
| 88 | 6 | 132309645 | 132314155 | 1490 |
| 89 | 11 | 5367204 | 5368185 | 390059 |
| 90 | 11 | 4781238 | 4782423 | 119695 |
| 91 | 13 | 112808105 | 112822346 | 2155 |
| 92 | 19 | 59158105 | 59177951 | 59283 |
| 93 | X | 153908257 | 153938385 | 65991 |
| 94 | 2 | 234316134 | 234317400 | 414061 |
| 95 | 21 | 32706622 | 32809568 | 59271 |
| 96 | 11 | 130745778 | 131710752 | 50863 |
| 97 | 21 | 32866419 | 32870062 | 55264 |
| 98 | 11 | 5203270 | 5204877 | 3043 |
| 99 | 11 | 5246158 | 5483410 | 3046 |
| 100 | 5 | 131466369 | 131511544 | 645029 |
| 101 | 11 | 5129236 | 5130175 | 23538 |
| 102 | 11 | 5714253 | 5716328 | 387748 |
| 103 | 21 | 33084854 | 33107868 | 56245 |
| 104 | 20 | 33720024 | 33750688 | 9054 |
| 105 | 5 | 132225179 | 132228124 | 2661 |
| 106 | 22 | 30845512 | 30846923 | 646580 |
| 107 | 11 | 5573934 | 5590217 | 117854 |
| 108 | 11 | 116196627 | 116199221 | 337 |
| 109 | 11 | 5210634 | 5212434 | 3045 |
| 110 | 9 | 130978873 | 131012683 | 389792 |
| 111 | 6 | 41714230 | 41729959 | 4188 |
| 112 | 11 | 5098469 | 5099384 | 390053 |
| 113 | 7 | 27134534 | 27136924 | 3201 |
| 114 | 1 | 149851285 | 149938183 | 81609 |
| 115 | 1 | 149851164 | 149933599 | 81609 |
| 116 | X | 152799320 | 152807619 | 643736 |
| 117 | 21 | 32870732 | 32879687 | 140290 |
| 118 | 18 | 59455481 | 59462470 | 6318 |
| 119 | 21 | 33066281 | 33093160 | 54067 |
| 120 | 11 | 5485107 | 5487744 | 50613 |

Continued on next page

**Table A.1 – continued from previous page**

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 121 | 19 | 59705824 | 59713709 | 3904 |
| 122 | 21 | 33779706 | 33785650 | 54943 |
| 123 | 5 | 131612501 | 131658907 | BC030525 |
| 124 | 19 | 59510164 | 59516221 | 353514 |
| 125 | 2 | 234209886 | 234343242 | 54575 |
| 126 | 18 | 59455932 | 59479430 | AF428135 |
| 127 | 5 | 132111041 | 132118263 | 645121 |
| 128 | 1 | 149750499 | 149777792 | 57530 |
| 129 | 13 | 112349358 | 112386812 | 400165 |
| 130 | 9 | 130896893 | 130912904 | 1384 |
| 131 | 7 | 127020924 | 127029079 | 29999 |
| 132 | 5 | 131658043 | 131707798 | 6583 |
| 133 | 18 | 59473411 | 59480098 | 6317 |
| 134 | 11 | 5641363 | 5662869 | 85363 |
| 135 | X | 153943079 | 153952830 | 4515 |
| 136 | 2 | 234333657 | 234346684 | 54658 |
| 137 | 7 | 27106497 | 27108919 | 3199 |
| 138 | X | 153254304 | 153256200 | AK125630 |
| 139 | 21 | 39699654 | 39739529 | 150082 |
| 140 | 11 | 64079673 | 64095575 | 55867 |
| 141 | 7 | 27151640 | 27153893 | 3203 |
| 142 | 7 | 27191681 | 27198951 | 646692 |
| 143 | 11 | 63934128 | 63944265 | 644541 |
| 144 | 6 | 41812427 | 41823099 | 5225 |
| 145 | 5 | 131621285 | 131637046 | 8572 |
| 146 | 7 | 27160814 | 27162821 | 3204 |
| 147 | 21 | 39739666 | 39809303 | 6450 |
| 148 | X | 122923269 | 123064027 | 10735 |
| 149 | 9 | 131138623 | 131140395 | AK092192 |
| 150 | X | 153952903 | 154004543 | 79184 |
| 151 | 11 | 5574461 | 5622204 | 445372 |
| 152 | 21 | 32922943 | 33022148 | 8867 |
| 153 | 21 | 33782367 | 33785893 | 54943 |
| 154 | 2 | 234490781 | 234592905 | 79054 |
| 155 | 11 | 2118322 | 2126470 | 51214 |
| 156 | 5 | 56240856 | 56248767 | 133383 |
| 157 | 13 | 112670814 | 112800864 | 23263 |
| 158 | 2 | 234624084 | 234650515 | 6694 |
| 159 | X | 122922235 | 123063026 | 10735 |
| 160 | 7 | 125865894 | 126670548 | 2918 |
| 161 | 7 | 115952074 | 115988466 | 857 |
| 162 | 21 | 32869022 | 32870472 | 55264 |

Continued on next page

**Table A.1 – continued from previous page**

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 163 | 7 | 27187653 | 27191355 | 3207 |
| 164 | 18 | 59528407 | 59541613 | 89778 |
| 165 | 7 | 27168581 | 27171674 | 3205 |
| 166 | 11 | 63973121 | 63975702 | 439914 |
| 167 | 7 | 117137940 | 117300797 | 83992 |
| 168 | 21 | 33936653 | 34183479 | 6453 |
| 169 | 7 | 116790511 | 116854779 | 136991 |
| 170 | 18 | 23784932 | 24011189 | 1000 |
| 171 | 7 | 116704517 | 116750579 | 7472 |
| 172 | 21 | 33320108 | 33323370 | 10215 |
| 173 | 22 | 30659507 | 30671336 | 25775 |
| 174 | 15 | 41652602 | 41769512 | 9677 |
| 175 | 18 | 59593623 | 59623592 | 8710 |
| 176 | 11 | 116165295 | 116167794 | 116519 |
| 177 | 7 | 113842511 | 114117391 | 93986 |
| 178 | 7 | 113842287 | 114117218 | 93986 |
| 179 | 7 | 27147520 | 27149812 | 3202 |
| 180 | 6 | 108722790 | 108950951 | 246269 |
| 181 | 21 | 34243099 | 34258130 | 400863 |
| 182 | 11 | 2273445 | 2279866 | 29125 |
| 183 | 7 | 116907252 | 117095951 | 1080 |
| 184 | 11 | 2106925 | 2109541 | 492304 |
| 185 | 7 | 89712444 | 89777638 | 79846 |
| 186 | 7 | 115926679 | 115935831 | 858 |
| 187 | 7 | 89678935 | 89704865 | 261729 |
| 188 | 7 | 89678993 | 89704927 | 261729 |
| 189 | 2 | 220016639 | 220039828 | 10290 |
| 190 | 7 | 89621624 | 89632077 | 26872 |
| 191 | 18 | 59705921 | 59722100 | 5055 |
| 192 | 19 | 59289813 | 59297806 | 126014 |
| 193 | 13 | 29674766 | 29779163 | 84056 |
| 194 | 2 | 219991342 | 219999705 | 1674 |
| 195 | 11 | 64114857 | 64126396 | 116085 |
| 196 | 7 | 27112333 | 27125739 | 3200 |
| 197 | 13 | 29680608 | 29779584 | 84056 |
| 198 | 2 | 220087135 | 220106998 | 55515 |
| 199 | 2 | 234351720 | 234406802 | 339766 |
| 200 | 21 | 33883516 | 33935936 | 9946 |
| 201 | 15 | 41906499 | 41946502 | 79968 |
| 202 | 11 | 2109739 | 2116400 | AK074614 |
| 203 | 7 | 27121491 | 27129028 | AK056230 |
| 204 | 5 | 132114415 | 132140966 | 23176 |

Continued on next page

**Table A.1 – continued from previous page**

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 205 | 12 | 38905085 | 39051870 | 120892 |
| 206 | 19 | 59289744 | 59295960 | 126014 |
| 207 | 13 | 112825145 | 112851842 | 2159 |
| 208 | 7 | 27203023 | 27206221 | 3209 |
| 209 | 18 | 59733724 | 59753456 | 5273 |
| 210 | 2 | 220087295 | 220111738 | 55515 |
| 211 | 11 | 116205833 | 116208997 | 345 |
| 212 | 11 | 64130221 | 64247236 | 9379 |
| 213 | 6 | 41845891 | 41855608 | 10817 |
| 214 | 11 | 2110355 | 2116780 | 3481 |
| 215 | X | 122821728 | 122875503 | 331 |
| 216 | 7 | 27099136 | 27102119 | 3198 |
| 217 | 7 | 114349444 | 114446492 | 29969 |
| 218 | 11 | 1953071 | 1956250 | AK126915 |
| 219 | 11 | 1972983 | 1975280 | 283120 |
| 220 | 14 | 98705376 | 98807575 | 64919 |
| 221 | 21 | 32687312 | 32688133 | 84996 |
| 222 | 10 | 55236344 | 55248144 | 387683 |
| 223 | 5 | 131733342 | 131759205 | 6584 |
| 224 | 2 | 220123699 | 220145134 | 23363 |
| 225 | 7 | 90729032 | 90731910 | 645794 |
| 226 | 11 | 116211678 | 116213548 | 335 |
| 227 | 20 | 33667222 | 33672379 | 6676 |
| 228 | 7 | 27176734 | 27180448 | 3206 |
| 229 | X | 153282772 | 153293621 | 1774 |
| 230 | 11 | 2279818 | 2296006 | 10077 |
| 231 | 16 | 48057 | 62591 | 64285 |
| 232 | 21 | 39479273 | 39607426 | 54014 |
| 233 | 21 | 39674139 | 39691496 | 7485 |
| 234 | 6 | 108469305 | 108502634 | 28962 |
| 235 | 11 | 2137586 | 2139015 | 3630 |
| 236 | 11 | 1817478 | 1819484 | 7136 |
| 237 | 19 | 59777070 | 59790833 | 11027 |
| 238 | 22 | 30650902 | 30652995 | AK123899 |
| 239 | 5 | 132059221 | 132101163 | 11127 |
| 240 | X | 153365249 | 153368126 | 8266 |
| 241 | X | 153368841 | 153372189 | 8273 |
| 242 | 7 | 116099694 | 116225676 | 4233 |
| 243 | 21 | 33364442 | 33366596 | 116448 |
| 244 | 5 | 131437383 | 131439758 | 1437 |
| 245 | 2 | 220200526 | 220214936 | 6508 |
| 246 | 12 | 38629561 | 38786156 | 114134 |

Continued on next page

**Table A.1 – continued from previous page**

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 247 | 19 | 59557944 | 59568280 | 3903 |
| 248 | 15 | 41815433 | 41825789 | 440278 |
| 249 | 11 | 64270605 | 64284763 | 5837 |
| 250 | 11 | 2141734 | 2149611 | 7054 |
| 251 | 5 | 132185910 | 132189901 | 134549 |
| 252 | 15 | 41612966 | 41669697 | 9677 |
| 253 | 11 | 1897511 | 1916512 | 7140 |
| 254 | 19 | 59064592 | 59071501 | 91663 |
| 255 | 16 | 142853 | 144504 | 3050 |
| 256 | 11 | 64348496 | 64368617 | 55561 |
| 257 | 19 | 59064504 | 59069685 | 91663 |
| 258 | 16 | 258310 | 265915 | 8786 |
| 259 | 21 | 33619083 | 33653999 | 3454 |
| 260 | 7 | 27248945 | 27252717 | 2128 |
| 261 | 2 | 220044627 | 220047344 | AK098307 |
| 262 | 19 | 59668021 | 59676234 | 148170 |
| 263 | 16 | 265611 | 277210 | 64714 |
| 264 | 11 | 1808892 | 1815326 | 90019 |
| 265 | 19 | 59796924 | 59804352 | 11024 |
| 266 | 5 | 56146021 | 56227730 | 4214 |
| 267 | X | 153293070 | 153303259 | 6901 |
| 268 | 21 | 33872080 | 33882884 | 29980 |
| 269 | 16 | 155972 | 156767 | 445449 |
| 270 | 2 | 220145197 | 220148671 | 3623 |
| 271 | 5 | 56251187 | 56283697 | 166968 |
| 272 | 7 | 90731718 | 90736068 | 8321 |
| 273 | 2 | 234438819 | 234441829 | 151507 |
| 274 | 16 | 261827 | 265981 | 8786 |
| 275 | 19 | 59618416 | 59639892 | 57348 |
| 276 | 9 | 130883226 | 130892538 | 57171 |
| 277 | 7 | 116447578 | 116657391 | 7982 |
| 278 | 6 | 74007762 | 74076659 | CR936715 |
| 279 | 16 | 162874 | 163708 | 3040 |
| 280 | 7 | 90063746 | 90674880 | 5218 |
| 281 | 5 | 132177177 | 132180377 | 134548 |
| 282 | 7 | 89870731 | 89883204 | 9069 |
| 283 | 11 | 2246303 | 2248758 | 430 |
| 284 | 9 | 130747629 | 130749833 | 22845 |
| 285 | 5 | 142130475 | 142586243 | 23092 |
| 286 | 5 | 142130155 | 142582945 | 23092 |
| 287 | 7 | 89813956 | 89858258 | 85865 |
| 288 | 16 | 166678 | 167520 | 3039 |

Continued on next page

**Table A.1 – continued from previous page**

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 289 | 19 | 59355657 | 59368664 | 147798 |
| 290 | 19 | 59355714 | 59368756 | 147798 |
| 291 | 19 | 59434640 | 59452868 | 79168 |
| 292 | 7 | 90176647 | 90677840 | 5218 |
| 293 | 7 | 116380616 | 116657313 | 7982 |
| 294 | 5 | 131920528 | 132007498 | 10111 |
| 295 | 19 | 59446172 | 59452939 | 10990 |
| 296 | 19 | 59412608 | 59438414 | 11025 |
| 297 | 21 | 33726662 | 33774120 | 757 |
| 298 | 1 | 149641823 | 149698556 | 23126 |
| 299 | X | 153359307 | 153360790 | 8270 |
| 300 | X | 153387699 | 153397567 | 60343 |
| 301 | 11 | 116124098 | 116148914 | 84811 |
| 302 | 2 | 118393639 | 118491788 | 54520 |
| 303 | 13 | 112392643 | 112589470 | 23250 |
| 304 | 15 | 41884024 | 41904243 | 4236 |
| 305 | X | 152940457 | 153016323 | 4204 |
| 306 | 9 | 130913064 | 130951044 | 5524 |
| 307 | 2 | 234410765 | 234427885 | 55355 |
| 308 | 21 | 39469253 | 39477310 | 8624 |
| 309 | 22 | 30480068 | 30633001 | 9681 |
| 310 | 1 | 149521414 | 149531005 | 57592 |
| 311 | 9 | 130978768 | 130980347 | 389792 |
| 312 | 7 | 116289798 | 116346549 | 830 |
| 313 | 19 | 59386005 | 59389333 | 79042 |
| 314 | 9 | 130839073 | 130874172 | 84895 |
| 315 | 7 | 127007917 | 127012890 | 79571 |
| 316 | 2 | 118389618 | 118390940 | 54520 |
| 317 | 9 | 130749797 | 130809195 | 23511 |
| 318 | 19 | 59491665 | 59496050 | 11026 |
| 319 | 2 | 220116988 | 220123561 | 130612 |
| 320 | 22 | 31110991 | 31113822 | 646621 |
| 321 | 21 | 33028083 | 33066040 | 94104 |
| 322 | 19 | 59664787 | 59666706 | 94059 |
| 323 | 11 | 64250958 | 64269504 | 10235 |
| 324 | 21 | 33021613 | 33022627 | 644266 |
| 325 | 21 | 39607756 | 39608756 | 257357 |
| 326 | X | 153412799 | 153428663 | 2539 |
| 327 | 21 | 33560541 | 33591390 | 3588 |
| 328 | 21 | 32895965 | 32906784 | 56683 |
| 329 | 12 | 38904567 | 38905165 | 642606 |
| 330 | 16 | 372247 | 382955 | 645631 |

Continued on next page

**Table A.1 – continued from previous page**

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 331 | X | 153310214 | 153318055 | 537 |
| 332 | 22 | 30670478 | 30683590 | 7533 |
| 333 | 16 | 277440 | 342465 | 8312 |
| 334 | 19 | 59351188 | 59355258 | 79165 |
| 335 | 20 | 33593191 | 33608819 | 51614 |
| 336 | 22 | 31201223 | 31224818 | 25793 |
| 337 | 16 | 170334 | 171178 | 3049 |
| 338 | 21 | 33798138 | 33836286 | 2618 |
| 339 | 20 | 33750740 | 33752294 | 140823 |
| 340 | 16 | 43016 | 47444 | 79622 |
| 341 | 6 | 108639409 | 108689156 | 8724 |
| 342 | X | 152823563 | 152825834 | 554 |
| 343 | X | 153339816 | 153355179 | 55558 |
| 344 | 5 | 132021763 | 132024700 | 3596 |
| 345 | 16 | 224801 | 258971 | 83986 |
| 346 | 22 | 31113568 | 31138235 | 51493 |
| 347 | 16 | 67017 | 75845 | 4350 |
| 348 | 22 | 30222260 | 30344534 | 9814 |
| 349 | 6 | 41856466 | 41865609 | 29964 |
| 350 | 5 | 132235912 | 132238286 | 116842 |
| 351 | 16 | 23876 | 26382 | 645582 |
| 352 | 1 | 149531036 | 149565348 | 5298 |
| 353 | 1 | 149437652 | 149488630 | 8394 |
| 354 | 1 | 149493820 | 149506560 | 5710 |
| 355 | 20 | 33330138 | 33336008 | 3692 |
| 356 | 1 | 149531231 | 149566511 | 5298 |
| 357 | 11 | 1730560 | 1741798 | 1509 |
| 358 | 19 | 59368920 | 59385478 | 79143 |
| 359 | 11 | 64313184 | 64327289 | 5871 |
| 360 | 16 | 415668 | 512482 | 9727 |
| 361 | 16 | 361859 | 371908 | 58986 |
| 362 | X | 153429255 | 153446455 | 8517 |
| 363 | 22 | 30165350 | 30215810 | 56478 |
| 364 | 7 | 126797588 | 126820003 | 168850 |
| 365 | 21 | 34197626 | 34210028 | 539 |
| 366 | 5 | 132037271 | 132046267 | 3565 |
| 367 | X | 122821265 | 122822820 | 643547 |
| 368 | 16 | 357396 | 360541 | 10573 |
| 369 | 16 | 387773 | 402487 | 26063 |
| 370 | 22 | 30402241 | 30438731 | 253143 |
| 371 | 16 | 356981 | 360226 | 10573 |
| 372 | 19 | 59866263 | 59873622 | 11006 |

Continued on next page

**Table A.1 – continued from previous page**

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 373 | 2 | 220111921 | 220116682 | 79586 |
| 374 | 6 | 108298214 | 108386086 | 11231 |
| 375 | 19 | 59536503 | 59542233 | 23547 |
| 376 | 11 | 64418594 | 64441239 | 23130 |
| 377 | 9 | 130810133 | 130830400 | 56904 |
| 378 | 19 | 59412548 | 59418709 | 11025 |
| 379 | 2 | 118310049 | 118312244 | 389024 |
| 380 | 2 | 118288724 | 118306423 | 8886 |
| 381 | 11 | 1925113 | 1934408 | 6150 |
| 382 | 7 | 115637816 | 115686073 | 26136 |
| 383 | 6 | 41622141 | 41678100 | 116113 |
| 384 | X | 153325695 | 153334051 | 9130 |
| 385 | X | 153260980 | 153263075 | 2010 |
| 386 | 7 | 127015694 | 127018989 | 381 |
| 387 | 2 | 220071856 | 220079955 | 29926 |
| 388 | 11 | 116154485 | 116163949 | 8882 |
| 389 | 21 | 39420865 | 39421836 | 391282 |
| 390 | 22 | 30411027 | 30439831 | 253143 |
| 391 | 21 | 39636110 | 39642917 | 3150 |
| 392 | 22 | 30160554 | 30164552 | AK127132 |
| 393 | 19 | 59470017 | 59476753 | 10288 |
| 394 | 20 | 33754944 | 33793607 | 9584 |
| 395 | X | 152848570 | 152853662 | 8260 |
| 396 | 19 | 59314702 | 59320534 | AK128544 |
| 397 | 20 | 33677379 | 33705245 | 8904 |
| 398 | 15 | 41825881 | 41852096 | 2923 |
| 399 | 16 | 178970 | 219450 | 55692 |
| 400 | 2 | 220170839 | 220189418 | 114790 |
| 401 | 20 | 33700290 | 33716252 | 10137 |
| 402 | 11 | 64302486 | 64303493 | 644613 |
| 403 | 5 | 131774571 | 131825958 | 441108 |
| 404 | 21 | 33837219 | 33871682 | 6651 |
| 405 | 16 | 387192 | 390755 | 4833 |
| 406 | 16 | 36999 | 43625 | 51728 |
| 407 | 15 | 41874456 | 41879547 | 619189 |
| 408 | 11 | 64327563 | 64334764 | 4221 |
| 409 | 21 | 33524100 | 33558697 | 3455 |
| 410 | 11 | 1830883 | 1870068 | 4046 |
| 411 | 19 | 59333260 | 59351239 | 4849 |
| 412 | 8 | 128875987 | 129182678 | 5820 |
| 413 | X | 153644343 | 153659154 | 1736 |
| 414 | X | 152929150 | 152938536 | 3654 |

Continued on next page

**Table A.1 – continued from previous page**

| # | chr | Start Site | End Site | Accession # |
|---|---|---|---|---|
| 415 | 15 | 41852089 | 41856794 | 80237 |
| 416 | 15 | 41879912 | 41882079 | 25764 |
| 417 | 18 | 59767573 | 59778624 | 284293 |
| 418 | 22 | 30345379 | 30388195 | 23761 |
| 419 | 7 | 27029881 | 27031053 | 402643 |
| 420 | 21 | 33697071 | 33731696 | 3460 |
| 421 | 16 | 4081 | 5847 | 375260 |
| 422 | X | 153318714 | 153325008 | 2664 |
| 423 | X | 152826026 | 152844908 | 393 |
| 424 | 15 | 41871843 | 41881362 | 25764 |
| 425 | X | 152866201 | 152883371 | 3054 |
| 426 | 6 | 41759693 | 41810776 | 7942 |
| 427 | 8 | 128816861 | 128821905 | M13930 |
| 428 | 22 | 30344476 | 30356810 | 23761 |
| 429 | 16 | 25950526 | 25951759 | 647915 |
| 430 | 11 | 64376783 | 64402767 | 10938 |
| 431 | 5 | 132230255 | 132231276 | 27089 |
| 432 | X | 153230158 | 153256123 | 2316 |
| 433 | 18 | 59788242 | 59807588 | 5271 |
| 434 | 19 | 59297971 | 59302080 | 4696 |
| 435 | 1 | 149638664 | 149641036 | 5692 |
| 436 | 8 | 128817497 | 128822856 | 4609 |
| 437 | X | 153279911 | 153283874 | 6134 |
| 438 | 1 | 149579739 | 149586393 | 5993 |
| 439 | 11 | 5667630 | 5688668 | 10346 |
| 440 | 19 | 59652207 | 59665006 | 114823 |
| 441 | 11 | 64288653 | 64302817 | 7536 |
| 442 | 19 | 59396537 | 59403327 | 6203 |
| 443 | 5 | 131846678 | 131854333 | 3659 |
| 444 | 22 | 30765440 | 30765968 | 402057 |
| 445 | 6 | 74283958 | 74287475 | 1915 |

# Appendix B

## Supplementary material for NanoString analysis

In this appendix lists of significantly affected genes is presented from various control and treatment groups. For more detals about these experiments, see chapter 7. For details about reading the tables, see section 7.4.2.1.

## B.1 Significance tables for various controls and treatment groups

Table B.1: List of significantly changed gene in the group: (1 - $H3122DMSO$, 2 - $H3122Criz$, 3 - $H3122NMS$), along with pairwise fold changes and significant change markers

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 1 | BMPR1A | 2.48 | 1.86 | 2.79 | 1.3 | -1.1 | -1.5 | 001 |
| 2 | BMPR2 | 2.48 | 1.86 | 2.79 | 1.3 | -1.1 | -1.5 | 001 |
| 3 | MAP2K2 | 6.51 | 5.89 | 7.13 | 1.1 | -1.1 | -1.2 | 001 |
| 4 | MAP4K4 | 6.51 | 5.89 | 6.82 | 1.1 | -1. | -1.2 | 001 |
| 5 | RIOK3 | 2.48 | 1.86 | 2.79 | 1.3 | -1.1 | -1.5 | 001 |
| 6 | SMG1 | 3.41 | 2.79 | 3.72 | 1.2 | -1.1 | -1.3 | 001 |
| 7 | CLK1 | 0.62 | 1.24 | 1.55 | -2. | -2.5 | -1.2 | 010 |
| 8 | CLK4 | 0.31 | 0.62 | 1.24 | -2. | -4. | -2. | 010 |
| 9 | PDIK1L | 1.55 | 1.86 | 2.48 | -1.2 | -1.6 | -1.3 | 010 |
| 10 | AURKB | 4.65 | 4.65 | 6.2 | -1. | -1.3 | -1.3 | 011 |
| 11 | IRAK1 | 5.89 | 5.27 | 7.13 | 1.1 | -1.2 | -1.4 | 011 |
| 12 | MELK | 6.51 | 6.2 | 7.44 | 1. | -1.1 | -1.2 | 011 |
| 13 | NEK7 | 4.34 | 4.34 | 5.27 | -1. | -1.2 | -1.2 | 011 |
| 14 | PIM1 | 0.93 | 1.55 | 2.48 | -1.7 | -2.7 | -1.6 | 011 |
| 15 | PLK1 | 5.89 | 5.58 | 7.13 | 1.1 | -1.2 | -1.3 | 011 |
| 16 | PRKDC | 24.49 | 23.87 | 31.62 | 1. | -1.3 | -1.3 | 011 |
| 17 | RIPK4 | 2.79 | 2.48 | 3.72 | 1.1 | -1.3 | -1.5 | 011 |
| 18 | TTK | 6.2 | 6.51 | 8.68 | -1. | -1.4 | -1.3 | 011 |
| 19 | WNK1 | 4.34 | 3.72 | 5.27 | 1.2 | -1.2 | -1.4 | 011 |
| 20 | ACVR1 | 2.79 | 1.86 | 2.48 | 1.5 | 1.1 | -1.3 | 100 |

Continued on next page

**Table B.1 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 21 | EGFR | 2.79 | 1.86 | 2.48 | 1.5 | 1.1 | -1.3 | 100 |
| 22 | EPHA4 | 2.79 | 1.86 | 2.17 | 1.5 | 1.3 | -1.2 | 100 |
| 23 | PAK1 | 4.34 | 3.41 | 3.72 | 1.3 | 1.2 | -1.1 | 100 |
| 24 | SCYL2 | 3.72 | 2.79 | 3.1 | 1.3 | 1.2 | -1.1 | 100 |
| 25 | STK40 | 1.55 | 0.62 | 0.93 | 2.5 | 1.7 | -1.5 | 100 |
| 26 | TP53RK | 2.17 | 1.24 | 1.55 | 1.8 | 1.4 | -1.2 | 100 |
| 27 | UHMK1 | 5.27 | 4.34 | 4.65 | 1.2 | 1.1 | -1.1 | 100 |
| 28 | YES1 | 5.27 | 4.34 | 4.96 | 1.2 | 1.1 | -1.1 | 100 |
| 29 | GSK3B | 6.82 | 5.58 | 6.51 | 1.2 | 1. | -1.2 | 101 |
| 30 | MAPK6 | 7.13 | 5.58 | 6.82 | 1.3 | 1. | -1.2 | 101 |
| 31 | MST4 | 5.27 | 4.03 | 4.96 | 1.3 | 1.1 | -1.2 | 101 |
| 32 | NRBP1 | 5.27 | 4.03 | 5.27 | 1.3 | -1. | -1.3 | 101 |
| 33 | PRKAA1 | 12.71 | 11.16 | 12.09 | 1.1 | 1.1 | -1.1 | 101 |
| 34 | PRKACA | 7.13 | 6.2 | 7.13 | 1.1 | -1. | -1.1 | 101 |
| 35 | PRPF4B | 5.58 | 4.03 | 5.27 | 1.4 | 1.1 | -1.3 | 101 |
| 36 | PTK2 | 13.64 | 12.09 | 13.02 | 1.1 | 1. | -1.1 | 101 |
| 37 | ROCK2 | 5.58 | 4.34 | 5.27 | 1.3 | 1.1 | -1.2 | 101 |
| 38 | SGK | 2.79 | 0.93 | 2.17 | 3. | 1.3 | -2.3 | 101 |
| 39 | SNF1LK | 2.79 | 1.55 | 3.1 | 1.8 | -1.1 | -2. | 101 |
| 40 | STK38L | 2.79 | 1.55 | 3.1 | 1.8 | -1.1 | -2. | 101 |
| 41 | CSNK1D | 17.05 | 15.81 | 15.19 | 1.1 | 1.1 | 1. | 110 |
| 42 | DAPK3 | 3.41 | 2.17 | 2.48 | 1.6 | 1.4 | -1.1 | 110 |
| 43 | EPHA2 | 8.06 | 1.24 | 1.55 | 6.5 | 5.2 | -1.2 | 110 |
| 44 | PLK2 | 18.6 | 4.65 | 4.96 | 4. | 3.8 | -1.1 | 110 |
| 45 | PLK3 | 1.55 | 0.31 | 0.31 | 5. | 5. | -1. | 110 |
| 46 | PRKD1 | 18.6 | 16.43 | 16.74 | 1.1 | 1.1 | -1. | 110 |
| 47 | SRPK1 | 5.27 | 4.03 | 4.34 | 1.3 | 1.2 | -1.1 | 110 |
| 48 | STK24 | 6.2 | 4.34 | 4.96 | 1.4 | 1.2 | -1.1 | 110 |
| 49 | AXL | 3.41 | 2.17 | 6.2 | 1.6 | -1.8 | -2.9 | 111 |
| 50 | BUB1B | 6.51 | 5.58 | 8.37 | 1.2 | -1.3 | -1.5 | 111 |
| 51 | CDC2 | 29.45 | 31. | 33.79 | -1.1 | -1.1 | -1.1 | 111 |
| 52 | CDK4 | 10.54 | 9.61 | 11.78 | 1.1 | -1.1 | -1.2 | 111 |
| 53 | CPNE3 | 7.13 | 4.96 | 6.2 | 1.4 | 1.1 | -1.2 | 111 |
| 54 | MET | 12.71 | 9.92 | 13.95 | 1.3 | -1.1 | -1.4 | 111 |
| 55 | STK17A | 13.64 | 6.2 | 9.92 | 2.2 | 1.4 | -1.6 | 111 |
| 56 | STK39 | 10.23 | 8.68 | 7.44 | 1.2 | 1.4 | 1.2 | 111 |
| 57 | TRIB1 | 8.06 | 3.72 | 6.51 | 2.2 | 1.2 | -1.8 | 111 |
| 58 | TRIO | 8.37 | 7.44 | 9.61 | 1.1 | -1.1 | -1.3 | 111 |

Table B.2: List of significantly changed gene in the group: (1 - $H2228DMSO$, 2 - $H2228Criz$, 3 - $H2228NMS$), along with pairwise fold changes and significant change markers

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 1 | BCR | 2.17 | 1.55 | 2.79 | 1.4 | -1.3 | -1.8 | 001 |
| 2 | CAMK2G | 2.17 | 1.55 | 2.48 | 1.4 | -1.1 | -1.6 | 001 |
| 3 | CAMKK2 | 2.17 | 1.55 | 2.79 | 1.4 | -1.3 | -1.8 | 001 |
| 4 | CDC2 | 23.87 | 24.49 | 23.25 | -1. | 1. | 1.1 | 001 |
| 5 | CDK7 | 2.17 | 1.86 | 2.79 | 1.2 | -1.3 | -1.5 | 001 |
| 6 | CDKL1 | 2.79 | 2.48 | 3.41 | 1.1 | -1.2 | -1.4 | 001 |
| 7 | COL4A3BP | 3.1 | 2.48 | 3.41 | 1.2 | -1.1 | -1.4 | 001 |
| 8 | CSNK1G1 | 1.55 | 1.24 | 2.17 | 1.2 | -1.4 | -1.8 | 001 |
| 9 | EIF2AK2 | 8.37 | 7.75 | 8.68 | 1.1 | -1. | -1.1 | 001 |
| 10 | ERN1 | 1.24 | 0.62 | 1.86 | 2. | -1.5 | -3. | 001 |
| 11 | FLJ13149 | 2.48 | 1.86 | 3.1 | 1.3 | -1.2 | -1.7 | 001 |
| 12 | HIPK1 | 0.93 | 0.62 | 1.55 | 1.5 | -1.7 | -2.5 | 001 |
| 13 | HUS1 | 5.58 | 4.96 | 6.2 | 1.1 | -1.1 | -1.2 | 001 |
| 14 | MAP2K6 | 0.62 | 0.31 | 1.24 | 2. | -2. | -4. | 001 |
| 15 | MAP3K2 | 4.65 | 4.34 | 5.27 | 1.1 | -1.1 | -1.2 | 001 |
| 16 | MAP3K7 | 3.72 | 3.1 | 4.03 | 1.2 | -1.1 | -1.3 | 001 |
| 17 | MAPK8 | 3.1 | 2.48 | 3.72 | 1.2 | -1.2 | -1.5 | 001 |
| 18 | MST1R | 3.1 | 2.48 | 3.41 | 1.2 | -1.1 | -1.4 | 001 |
| 19 | PKN2 | 2.79 | 2.17 | 3.1 | 1.3 | -1.1 | -1.4 | 001 |
| 20 | SLK | 5.27 | 4.65 | 5.89 | 1.1 | -1.1 | -1.3 | 001 |
| 21 | SNRK | 0.93 | 0.62 | 1.55 | 1.5 | -1.7 | -2.5 | 001 |
| 22 | SRC | 0.93 | 0.62 | 1.55 | 1.5 | -1.7 | -2.5 | 001 |
| 23 | STK32A | 2.48 | 1.86 | 3.1 | 1.3 | -1.2 | -1.7 | 001 |
| 24 | TAF1 | 1.55 | 1.24 | 2.17 | 1.2 | -1.4 | -1.8 | 001 |
| 25 | TAF1L | 1.24 | 0.93 | 1.86 | 1.3 | -1.5 | -2. | 001 |
| 26 | TAOK1 | 4.34 | 3.72 | 4.65 | 1.2 | -1.1 | -1.2 | 001 |
| 27 | TBRG4 | 2.79 | 2.17 | 3.1 | 1.3 | -1.1 | -1.4 | 001 |
| 28 | ULK1 | 1.24 | 0.93 | 1.86 | 1.3 | -1.5 | -2. | 001 |
| 29 | ULK3 | 1.86 | 1.24 | 2.17 | 1.5 | -1.2 | -1.8 | 001 |
| 30 | ZAK | 1.55 | 0.93 | 1.86 | 1.7 | -1.2 | -2. | 001 |
| 31 | CLK1 | 0.62 | 1.24 | 1.55 | -2. | -2.5 | -1.2 | 010 |
| 32 | RAF1 | 6.51 | 5.89 | 7.44 | 1.1 | -1.1 | -1.3 | 011 |
| 33 | RPS6KB1 | 2.17 | 2.17 | 3.1 | -1. | -1.4 | -1.4 | 011 |
| 34 | SMG1 | 4.65 | 4.34 | 5.89 | 1.1 | -1.3 | -1.4 | 011 |
| 35 | STK24 | 6.82 | 6.51 | 7.75 | 1. | -1.1 | -1.2 | 011 |
| 36 | TRIO | 5.58 | 4.96 | 6.51 | 1.1 | -1.2 | -1.3 | 011 |
| 37 | ADCK2 | 2.48 | 1.55 | 2.17 | 1.6 | 1.1 | -1.4 | 100 |
| 38 | BMPR1A | 5.27 | 4.03 | 4.65 | 1.3 | 1.1 | -1.2 | 100 |
| 39 | CHUK | 5.27 | 4.34 | 4.65 | 1.2 | 1.1 | -1.1 | 100 |

Continued on next page

**Table B.2 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 40 | ILK | 3.41 | 2.48 | 3.1 | 1.4 | 1.1 | -1.2 | 100 |
| 41 | MAP2K3 | 1.55 | 0.62 | 1.24 | 2.5 | 1.2 | -2. | 100 |
| 42 | PKMYT1 | 0.31 | 1.24 | 0.93 | -4. | -3. | 1.3 | 100 |
| 43 | STK16 | 1.86 | 0.93 | 1.55 | 2. | 1.2 | -1.7 | 100 |
| 44 | WNK1 | 13.02 | 11.78 | 12.4 | 1.1 | 1. | -1.1 | 100 |
| 45 | ABL1 | 3.41 | 2.48 | 3.72 | 1.4 | -1.1 | -1.5 | 101 |
| 46 | ABL2 | 2.79 | 1.55 | 2.79 | 1.8 | -1. | -1.8 | 101 |
| 47 | ADRBK1 | 7.75 | 5.27 | 7.44 | 1.5 | 1. | -1.4 | 101 |
| 48 | AKT2 | 7.13 | 5.58 | 7.44 | 1.3 | -1. | -1.3 | 101 |
| 49 | ALPK1 | 2.17 | 0.93 | 2.48 | 2.3 | -1.1 | -2.7 | 101 |
| 50 | ARAF | 4.65 | 3.72 | 5.27 | 1.3 | -1.1 | -1.4 | 101 |
| 51 | AXL | 26.97 | 32.24 | 26.35 | -1.2 | 1. | 1.2 | 101 |
| 52 | BCKDK | 2.79 | 1.86 | 3.1 | 1.5 | -1.1 | -1.7 | 101 |
| 53 | BRD2 | 5.89 | 4.34 | 6.51 | 1.4 | -1.1 | -1.5 | 101 |
| 54 | BUB1B | 5.27 | 6.2 | 5.27 | -1.2 | -1. | 1.2 | 101 |
| 55 | CAMK1 | 1.24 | 0.31 | 1.24 | 4. | -1. | -4. | 101 |
| 56 | CAMK2D | 8.37 | 4.96 | 8.68 | 1.7 | -1. | -1.8 | 101 |
| 57 | CDC2L5 | 4.03 | 3.1 | 4.34 | 1.3 | -1.1 | -1.4 | 101 |
| 58 | CDK10 | 5.58 | 3.72 | 5.58 | 1.5 | -1. | -1.5 | 101 |
| 59 | CDK4 | 33.17 | 27.9 | 33.79 | 1.2 | -1. | -1.2 | 101 |
| 60 | CDK9 | 3.41 | 2.48 | 3.72 | 1.4 | -1.1 | -1.5 | 101 |
| 61 | CPNE3 | 6.82 | 4.96 | 6.82 | 1.4 | -1. | -1.4 | 101 |
| 62 | CSNK1E | 6.82 | 5.27 | 7.13 | 1.3 | -1. | -1.4 | 101 |
| 63 | CSNK1G3 | 5.27 | 4.34 | 5.58 | 1.2 | -1.1 | -1.3 | 101 |
| 64 | CSNK2A1 | 5.89 | 4.65 | 5.58 | 1.3 | 1.1 | -1.2 | 101 |
| 65 | DAPK3 | 5.89 | 3.41 | 5.89 | 1.7 | -1. | -1.7 | 101 |
| 66 | DDR1 | 2.48 | 1.55 | 2.79 | 1.6 | -1.1 | -1.8 | 101 |
| 67 | DYRK3 | 3.72 | 2.48 | 4.03 | 1.5 | -1.1 | -1.6 | 101 |
| 68 | EGFR | 8.37 | 6.2 | 8.06 | 1.3 | 1. | -1.3 | 101 |
| 69 | EIF2AK3 | 2.17 | 0.93 | 2.17 | 2.3 | -1. | -2.3 | 101 |
| 70 | EIF2AK4 | 7.13 | 5.89 | 7.13 | 1.2 | -1. | -1.2 | 101 |
| 71 | HIPK3 | 2.79 | 1.55 | 2.79 | 1.8 | -1. | -1.8 | 101 |
| 72 | IRAK2 | 3.72 | 2.17 | 3.72 | 1.7 | -1. | -1.7 | 101 |
| 73 | LATS1 | 3.1 | 1.86 | 3.41 | 1.7 | -1.1 | -1.8 | 101 |
| 74 | LATS2 | 2.17 | 0.93 | 2.17 | 2.3 | -1. | -2.3 | 101 |
| 75 | LYN | 5.27 | 2.79 | 5.27 | 1.9 | -1. | -1.9 | 101 |
| 76 | MAP2K2 | 26.35 | 20.15 | 26.97 | 1.3 | -1. | -1.3 | 101 |
| 77 | MAP3K13 | 1.24 | 0.31 | 1.55 | 4. | -1.2 | -5. | 101 |
| 78 | MAP4K5 | 10.85 | 8.06 | 10.54 | 1.3 | 1. | -1.3 | 101 |
| 79 | MAPK1 | 6.82 | 5.58 | 6.51 | 1.2 | 1. | -1.2 | 101 |
| 80 | MAPK13 | 1.55 | 0.62 | 1.55 | 2.5 | -1. | -2.5 | 101 |
| 81 | MAPK3 | 5.27 | 3.1 | 5.89 | 1.7 | -1.1 | -1.9 | 101 |

Continued on next page

**Table B.2 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 82 | MAPK9 | 5.89 | 4.96 | 5.89 | 1.2 | -1. | -1.2 | 101 |
| 83 | MAPKAPK2 | 6.2 | 4.65 | 5.89 | 1.3 | 1.1 | -1.3 | 101 |
| 84 | MAPKAPK5 | 3.72 | 2.79 | 4.03 | 1.3 | -1.1 | -1.4 | 101 |
| 85 | MARK4 | 2.48 | 1.55 | 2.79 | 1.6 | -1.1 | -1.8 | 101 |
| 86 | MELK | 6.82 | 9.3 | 7.44 | -1.4 | -1.1 | 1.3 | 101 |
| 87 | MKNK2 | 7.44 | 4.34 | 7.75 | 1.7 | -1. | -1.8 | 101 |
| 88 | MST4 | 4.65 | 3.72 | 5.27 | 1.3 | -1.1 | -1.4 | 101 |
| 89 | NRBP1 | 8.68 | 6.2 | 8.68 | 1.4 | -1. | -1.4 | 101 |
| 90 | OXSR1 | 4.34 | 3.41 | 4.34 | 1.3 | -1. | -1.3 | 101 |
| 91 | PAK1 | 7.13 | 4.96 | 7.44 | 1.4 | -1. | -1.5 | 101 |
| 92 | PAK2 | 8.06 | 6.51 | 8.06 | 1.2 | -1. | -1.2 | 101 |
| 93 | PBK | 7.44 | 9.61 | 8.06 | -1.3 | -1.1 | 1.2 | 101 |
| 94 | PCTK1 | 8.68 | 6.2 | 8.99 | 1.4 | -1. | -1.4 | 101 |
| 95 | PDK3 | 3.1 | 1.86 | 3.1 | 1.7 | -1. | -1.7 | 101 |
| 96 | PDPK1 | 3.41 | 2.48 | 3.72 | 1.4 | -1.1 | -1.5 | 101 |
| 97 | PIM2 | 2.48 | 1.24 | 2.17 | 2. | 1.1 | -1.8 | 101 |
| 98 | PIM3 | 2.17 | 0.93 | 1.86 | 2.3 | 1.2 | -2. | 101 |
| 99 | PKN1 | 9.61 | 6.2 | 8.99 | 1.5 | 1.1 | -1.4 | 101 |
| 100 | PLK1 | 10.54 | 14.88 | 10.23 | -1.4 | 1. | 1.5 | 101 |
| 101 | PRKAA1 | 5.58 | 4.34 | 5.89 | 1.3 | -1.1 | -1.4 | 101 |
| 102 | PRKACA | 21.39 | 18.29 | 21.7 | 1.2 | -1. | -1.2 | 101 |
| 103 | PRKCI | 6.82 | 4.65 | 7.13 | 1.5 | -1. | -1.5 | 101 |
| 104 | PRPF4B | 12.09 | 10.23 | 12.4 | 1.2 | -1. | -1.2 | 101 |
| 105 | PTK2 | 10.54 | 7.13 | 9.92 | 1.5 | 1.1 | -1.4 | 101 |
| 106 | RIOK2 | 5.27 | 4.34 | 5.89 | 1.2 | -1.1 | -1.4 | 101 |
| 107 | RIPK1 | 5.27 | 3.72 | 5.58 | 1.4 | -1.1 | -1.5 | 101 |
| 108 | RIPK2 | 5.58 | 4.03 | 5.27 | 1.4 | 1.1 | -1.3 | 101 |
| 109 | RPS6KA1 | 3.41 | 2.48 | 3.41 | 1.4 | -1. | -1.4 | 101 |
| 110 | RPS6KA3 | 2.17 | 1.24 | 2.79 | 1.8 | -1.3 | -2.2 | 101 |
| 111 | SCYL1 | 4.34 | 3.1 | 4.34 | 1.4 | -1. | -1.4 | 101 |
| 112 | SCYL2 | 5.58 | 4.65 | 5.89 | 1.2 | -1.1 | -1.3 | 101 |
| 113 | SNF1LK2 | 6.2 | 4.34 | 6.51 | 1.4 | -1. | -1.5 | 101 |
| 114 | SRPK1 | 12.09 | 10.54 | 11.78 | 1.1 | 1. | -1.1 | 101 |
| 115 | SRPK2 | 3.41 | 2.48 | 3.41 | 1.4 | -1. | -1.4 | 101 |
| 116 | STK17B | 12.09 | 5.58 | 11.47 | 2.2 | 1.1 | -2.1 | 101 |
| 117 | STK38 | 4.65 | 3.72 | 4.65 | 1.3 | -1. | -1.2 | 101 |
| 118 | STK40 | 3.72 | 2.79 | 4.03 | 1.3 | -1.1 | -1.4 | 101 |
| 119 | TAOK3 | 5.58 | 3.72 | 5.89 | 1.5 | -1.1 | -1.6 | 101 |
| 120 | TBK1 | 4.03 | 2.79 | 4.65 | 1.4 | -1.2 | -1.7 | 101 |
| 121 | TLK1 | 4.34 | 3.41 | 4.34 | 1.3 | -1. | -1.3 | 101 |
| 122 | TRIB2 | 2.79 | 1.24 | 2.48 | 2.2 | 1.1 | -2. | 101 |
| 123 | UHMK1 | 6.82 | 5.89 | 7.13 | 1.2 | -1. | -1.2 | 101 |

Continued on next page

**Table B.2 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 124 | YES1 | 8.68 | 7.13 | 8.37 | 1.2 | 1. | -1.2 | 101 |
| 125 | CCL2 | 4.34 | 1.55 | 1.86 | 2.8 | 2.3 | -1.2 | 110 |
| 126 | MAP3K14 | 2.17 | 0.93 | 1.24 | 2.3 | 1.8 | -1.3 | 110 |
| 127 | MAP4K4 | 13.33 | 14.57 | 14.57 | -1.1 | -1.1 | -1. | 110 |
| 128 | MAPK14 | 10.54 | 8.68 | 9.3 | 1.2 | 1.1 | -1.1 | 110 |
| 129 | PRKDC | 11.47 | 13.33 | 13.33 | -1.2 | -1.2 | -1. | 110 |
| 130 | SGK | 4.34 | 2.48 | 2.17 | 1.8 | 2. | 1.1 | 110 |
| 131 | ACVR1B | 1.86 | 0.93 | 2.79 | 2. | -1.5 | -3. | 111 |
| 132 | CSNK1A1 | 11.47 | 8.06 | 12.4 | 1.4 | -1.1 | -1.5 | 111 |
| 133 | CSNK1D | 26.66 | 15.5 | 25.73 | 1.7 | 1. | -1.7 | 111 |
| 134 | EIF2AK1 | 11.47 | 8.37 | 12.4 | 1.4 | -1.1 | -1.5 | 111 |
| 135 | EPHA2 | 3.72 | 1.24 | 2.17 | 3. | 1.7 | -1.8 | 111 |
| 136 | GRK6 | 5.58 | 3.1 | 6.51 | 1.8 | -1.2 | -2.1 | 111 |
| 137 | GSK3B | 9.61 | 7.75 | 10.54 | 1.2 | -1.1 | -1.4 | 111 |
| 138 | MAP2K1 | 15.19 | 13.02 | 14.26 | 1.2 | 1.1 | -1.1 | 111 |
| 139 | MAP3K5 | 1.86 | 0.93 | 2.79 | 2. | -1.5 | -3. | 111 |
| 140 | MAPK6 | 15.5 | 10.85 | 14.26 | 1.4 | 1.1 | -1.3 | 111 |
| 141 | MET | 26.35 | 24.8 | 27.28 | 1.1 | -1. | -1.1 | 111 |
| 142 | MYLK | 5.27 | 4.34 | 6.2 | 1.2 | -1.2 | -1.4 | 111 |
| 143 | NEK7 | 6.82 | 5.58 | 8.06 | 1.2 | -1.2 | -1.4 | 111 |
| 144 | NUAK2 | 4.34 | 1.24 | 2.48 | 3.5 | 1.8 | -2. | 111 |
| 145 | PIM1 | 4.03 | 3.1 | 6.51 | 1.3 | -1.6 | -2.1 | 111 |
| 146 | PLK2 | 33.48 | 15.5 | 9.3 | 2.2 | 3.6 | 1.7 | 111 |
| 147 | RIPK4 | 6.82 | 5.89 | 8.06 | 1.2 | -1.2 | -1.4 | 111 |
| 148 | ROS1 | 5.58 | 2.79 | 6.82 | 2. | -1.2 | -2.4 | 111 |
| 149 | STK17A | 48.36 | 38.44 | 46.81 | 1.3 | 1. | -1.2 | 111 |
| 150 | STK39 | 5.58 | 4.65 | 6.51 | 1.2 | -1.2 | -1.4 | 111 |
| 151 | TGFBR2 | 45.57 | 27.9 | 52.08 | 1.6 | -1.1 | -1.9 | 111 |
| 152 | TRIB1 | 5.89 | 2.48 | 4.96 | 2.4 | 1.2 | -2. | 111 |
| 153 | TRIB3 | 6.51 | 4.03 | 8.06 | 1.6 | -1.2 | -2. | 111 |

Table B.3: List of significantly changed gene in the group: (1 - $BEAS2BPar$, 2 - $BEAS2BWT$, 3 - $BEAS2BKR$), along with pairwise fold changes and significant change markers

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 1 | CAMK2G | 3.1 | 2.48 | 3.72 | 1.2 | -1.2 | -1.5 | 001 |
| 2 | CAMKK2 | 2.79 | 2.17 | 3.41 | 1.3 | -1.2 | -1.6 | 001 |
| 3 | CDK7 | 2.79 | 2.17 | 3.1 | 1.3 | -1.1 | -1.4 | 001 |
| 4 | FLJ25006 | 0.31 | 0. | 0.93 | 100. | -3. | -300. | 001 |
| 5 | LYK5 | 0.93 | 0.31 | 1.24 | 3. | -1.3 | -4. | 001 |
| 6 | MAP3K9 | 0.62 | 0. | 0.93 | 200. | -1.5 | -300. | 001 |
| 7 | MERTK | 0.62 | 0. | 0.93 | 200. | -1.5 | -300. | 001 |
| 8 | MINK1 | 3.1 | 2.48 | 3.72 | 1.2 | -1.2 | -1.5 | 001 |
| 9 | NLK | 1.55 | 0.93 | 2.17 | 1.7 | -1.4 | -2.3 | 001 |
| 10 | NUAK2 | 2.79 | 2.17 | 3.1 | 1.3 | -1.1 | -1.4 | 001 |
| 11 | PDGFRB | 1.55 | 1.86 | 0.93 | -1.2 | 1.7 | 2. | 001 |
| 12 | STK10 | 1.24 | 0.93 | 1.86 | 1.3 | -1.5 | -2. | 001 |
| 13 | STK16 | 1.24 | 0.62 | 1.86 | 2. | -1.5 | -3. | 001 |
| 14 | STK33 | 0.62 | 0.31 | 1.24 | 2. | -2. | -4. | 001 |
| 15 | TNK1 | 0.93 | 0.31 | 1.55 | 3. | -1.7 | -5. | 001 |
| 16 | TNK2 | 0.62 | 0. | 0.93 | 200. | -1.5 | -300. | 001 |
| 17 | ADCK5 | 0.31 | 0. | 1.24 | 100. | -4. | -400. | 011 |
| 18 | ALPK1 | 0.93 | 0.31 | 2.17 | 3. | -2.3 | -7. | 011 |
| 19 | CDK6 | 1.86 | 1.86 | 7.44 | -1. | -4. | -4. | 011 |
| 20 | CDK9 | 3.41 | 2.79 | 7.13 | 1.2 | -2.1 | -2.6 | 011 |
| 21 | DAPK1 | 2.79 | 2.79 | 0.31 | -1. | 9. | 9. | 011 |
| 22 | DDR1 | 3.1 | 3.1 | 9.61 | -1. | -3.1 | -3.1 | 011 |
| 23 | DYRK3 | 1.55 | 0.93 | 2.79 | 1.7 | -1.8 | -3. | 011 |
| 24 | EPHA1 | 0. | 0. | 1.55 | -1. | -500. | -500. | 011 |
| 25 | EPHB2 | 0.62 | 0.31 | 2.17 | 2. | -3.5 | -7. | 011 |
| 26 | ERBB3 | 0.62 | 0.31 | 3.72 | 2. | -6. | -12. | 011 |
| 27 | MAPK13 | 3.41 | 4.03 | 14.57 | -1.2 | -4.3 | -3.6 | 011 |
| 28 | MAPK14 | 10.54 | 10.23 | 13.33 | 1. | -1.3 | -1.3 | 011 |
| 29 | MARK1 | 0.31 | 0. | 1.55 | 100. | -5. | -500. | 011 |
| 30 | MGC5297 | 1.55 | 1.86 | 3.41 | -1.2 | -2.2 | -1.8 | 011 |
| 31 | PAK6 | 0.62 | 0.31 | 2.48 | 2. | -4. | -8. | 011 |
| 32 | PRKCD | 2.48 | 2.17 | 4.34 | 1.1 | -1.8 | -2. | 011 |
| 33 | PRKCZ | 1.86 | 1.24 | 3.41 | 1.5 | -1.8 | -2.8 | 011 |
| 34 | PRKD1 | 4.34 | 4.34 | 1.24 | -1. | 3.5 | 3.5 | 011 |
| 35 | RAGE | 1.86 | 1.86 | 5.27 | -1. | -2.8 | -2.8 | 011 |
| 36 | STK17A | 3.72 | 4.03 | 9.61 | -1.1 | -2.6 | -2.4 | 011 |
| 37 | STK17B | 6.51 | 7.13 | 5.58 | -1.1 | 1.2 | 1.3 | 011 |
| 38 | TBRG4 | 2.48 | 2.48 | 3.72 | -1. | -1.5 | -1.5 | 011 |
| 39 | CDKL2 | 0.93 | 0. | 0.62 | 300. | 1.5 | -200. | 100 |

Continued on next page

**Table B.3 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 40 | IKBKB | 1.24 | 0.31 | 0.62 | 4. | 2. | -2. | 100 |
| 41 | IRAK3 | 1.24 | 0.31 | 0.62 | 4. | 2. | -2. | 100 |
| 42 | MAP3K1 | 2.17 | 1.24 | 1.55 | 1.8 | 1.4 | -1.2 | 100 |
| 43 | PLK4 | 1.24 | 0.31 | 0.93 | 4. | 1.3 | -3. | 100 |
| 44 | ACVR2A | 2.48 | 1.24 | 2.79 | 2. | -1.1 | -2.2 | 101 |
| 45 | AKT3 | 8.06 | 3.41 | 8.68 | 2.4 | -1.1 | -2.5 | 101 |
| 46 | CDC2L6 | 4.03 | 1.24 | 3.41 | 3.2 | 1.2 | -2.8 | 101 |
| 47 | CDC42BPA | 2.17 | 0.62 | 2.79 | 3.5 | -1.3 | -4.5 | 101 |
| 48 | CDKL1 | 1.24 | 0. | 0.93 | 400. | 1.3 | -300. | 101 |
| 49 | CDKL3 | 2.17 | 0.31 | 1.86 | 7. | 1.2 | -6. | 101 |
| 50 | CLK3 | 3.41 | 0.62 | 3.41 | 5.5 | -1. | -5.5 | 101 |
| 51 | EPHB4 | 1.55 | 0.62 | 2.17 | 2.5 | -1.4 | -3.5 | 101 |
| 52 | FGFRL1 | 0.93 | 0. | 1.24 | 300. | -1.3 | -400. | 101 |
| 53 | FYN | 7.13 | 1.24 | 7.75 | 5.8 | -1.1 | -6.2 | 101 |
| 54 | HIPK1 | 2.79 | 1.24 | 2.48 | 2.2 | 1.1 | -2. | 101 |
| 55 | MAP2K3 | 0.31 | 1.86 | 0.62 | -6. | -2. | 3. | 101 |
| 56 | MAP2K7 | 3.41 | 1.86 | 3.72 | 1.8 | -1.1 | -2. | 101 |
| 57 | MAP3K10 | 0.93 | 0. | 1.55 | 300. | -1.7 | -500. | 101 |
| 58 | MAP3K14 | 2.48 | 1.55 | 3.1 | 1.6 | -1.2 | -2. | 101 |
| 59 | MAP3K3 | 8.37 | 4.96 | 8.99 | 1.7 | -1.1 | -1.8 | 101 |
| 60 | MAP3K4 | 3.72 | 1.86 | 4.03 | 2. | -1.1 | -2.2 | 101 |
| 61 | MAPK12 | 2.17 | 0.62 | 1.55 | 3.5 | 1.4 | -2.5 | 101 |
| 62 | MAPK7 | 3.72 | 1.86 | 3.72 | 2. | -1. | -2. | 101 |
| 63 | MAST2 | 3.72 | 1.24 | 4.03 | 3. | -1.1 | -3.2 | 101 |
| 64 | MKNK2 | 8.37 | 4.03 | 8.68 | 2.1 | -1. | -2.2 | 101 |
| 65 | MLKL | 0.62 | 1.86 | 0.93 | -3. | -1.5 | 2. | 101 |
| 66 | NEK4 | 1.55 | 0.62 | 1.55 | 2.5 | -1. | -2.5 | 101 |
| 67 | NEK9 | 2.48 | 1.24 | 2.79 | 2. | -1.1 | -2.2 | 101 |
| 68 | NRBP1 | 19.53 | 11.16 | 20.15 | 1.8 | -1. | -1.8 | 101 |
| 69 | NRBP2 | 1.86 | 0.31 | 1.55 | 6. | 1.2 | -5. | 101 |
| 70 | PASK | 1.24 | 0. | 1.24 | 400. | -1. | -400. | 101 |
| 71 | PBK | 10.54 | 7.75 | 9.92 | 1.4 | 1.1 | -1.3 | 101 |
| 72 | PDPK1 | 6.51 | 2.48 | 5.89 | 2.6 | 1.1 | -2.4 | 101 |
| 73 | PHKG2 | 3.41 | 2.17 | 3.41 | 1.6 | -1. | -1.6 | 101 |
| 74 | PRKAA2 | 3.1 | 0.31 | 3.1 | 10. | -1. | -10. | 101 |
| 75 | PRKACB | 3.72 | 1.86 | 4.03 | 2. | -1.1 | -2.2 | 101 |
| 76 | PRKCE | 1.86 | 0.62 | 1.55 | 3. | 1.2 | -2.5 | 101 |
| 77 | PSKH1 | 1.55 | 0.31 | 1.86 | 5. | -1.2 | -6. | 101 |
| 78 | PXK | 2.48 | 0.62 | 1.86 | 4. | 1.3 | -3. | 101 |
| 79 | RIOK1 | 1.55 | 0.62 | 1.86 | 2.5 | -1.2 | -3. | 101 |
| 80 | RPS6KA1 | 9.3 | 8.06 | 9.61 | 1.2 | -1. | -1.2 | 101 |
| 81 | SCYL3 | 2.79 | 0.93 | 3.41 | 3. | -1.2 | -3.7 | 101 |

Continued on next page

**Table B.3 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 82 | SGK269 | 3.72 | 0.93 | 3.72 | 4. | -1. | -4. | 101 |
| 83 | SGK3 | 2.17 | 0.62 | 2.48 | 3.5 | -1.1 | -4. | 101 |
| 84 | STK19 | 1.24 | 0. | 1.24 | 400. | -1. | -400. | 101 |
| 85 | STK25 | 5.27 | 2.17 | 5.27 | 2.4 | -1. | -2.4 | 101 |
| 86 | STK3 | 2.17 | 0.31 | 2.48 | 7. | -1.1 | -8. | 101 |
| 87 | STK35 | 4.03 | 1.55 | 4.65 | 2.6 | -1.2 | -3. | 101 |
| 88 | STK38 | 6.51 | 3.72 | 6.82 | 1.8 | -1. | -1.8 | 101 |
| 89 | STK38L | 2.79 | 1.55 | 3.41 | 1.8 | -1.2 | -2.2 | 101 |
| 90 | STK4 | 4.34 | 1.55 | 3.72 | 2.8 | 1.2 | -2.4 | 101 |
| 91 | TBK1 | 6.2 | 3.72 | 6.82 | 1.7 | -1.1 | -1.8 | 101 |
| 92 | TESK1 | 0.93 | 0. | 1.24 | 300. | -1.3 | -400. | 101 |
| 93 | TGFBR1 | 2.17 | 0.31 | 2.17 | 7. | -1. | -7. | 101 |
| 94 | TYK2 | 1.24 | 0.31 | 1.24 | 4. | -1. | -4. | 101 |
| 95 | TYRO3 | 1.86 | 0.31 | 1.24 | 6. | 1.5 | -4. | 101 |
| 96 | ULK1 | 2.17 | 0.62 | 2.79 | 3.5 | -1.3 | -4.5 | 101 |
| 97 | ULK2 | 3.41 | 0. | 2.79 | 1100. | 1.2 | -900. | 101 |
| 98 | ZAK | 2.17 | 0.62 | 1.55 | 3.5 | 1.4 | -2.5 | 101 |
| 99 | AURKB | 11.78 | 13.02 | 13.64 | -1.1 | -1.2 | -1. | 110 |
| 100 | BCKDK | 4.03 | 5.27 | 5.58 | -1.3 | -1.4 | -1.1 | 110 |
| 101 | BUB1B | 9.3 | 10.23 | 10.85 | -1.1 | -1.2 | -1.1 | 110 |
| 102 | CAMK1D | 1.24 | 0.31 | 0.31 | 4. | 4. | -1. | 110 |
| 103 | CDK4 | 28.83 | 23.56 | 23.56 | 1.2 | 1.2 | -1. | 110 |
| 104 | PDK2 | 3.41 | 2.17 | 1.86 | 1.6 | 1.8 | 1.2 | 110 |
| 105 | PRKACA | 22.32 | 21.08 | 20.46 | 1.1 | 1.1 | 1. | 110 |
| 106 | ROR2 | 4.65 | 1.24 | 1.86 | 3.8 | 2.5 | -1.5 | 110 |
| 107 | SNF1LK | 1.86 | 3.1 | 2.79 | -1.7 | -1.5 | 1.1 | 110 |
| 108 | AAK1 | 8.37 | 2.48 | 12.71 | 3.4 | -1.5 | -5.1 | 111 |
| 109 | ABL1 | 3.72 | 1.86 | 5.58 | 2. | -1.5 | -3. | 111 |
| 110 | ABL2 | 4.03 | 0.93 | 6.51 | 4.3 | -1.6 | -7. | 111 |
| 111 | ACVR1 | 4.34 | 1.55 | 9.92 | 2.8 | -2.3 | -6.4 | 111 |
| 112 | ACVR1B | 3.41 | 1.24 | 5.89 | 2.8 | -1.7 | -4.8 | 111 |
| 113 | ADCK2 | 7.13 | 2.48 | 13.95 | 2.9 | -2. | -5.6 | 111 |
| 114 | ADRBK1 | 11.78 | 6.51 | 16.12 | 1.8 | -1.4 | -2.5 | 111 |
| 115 | AKT2 | 8.68 | 6.51 | 12.4 | 1.3 | -1.4 | -1.9 | 111 |
| 116 | ALS2CR2 | 2.79 | 0.62 | 1.55 | 4.5 | 1.8 | -2.5 | 111 |
| 117 | ARAF | 6.2 | 4.96 | 7.44 | 1.2 | -1.2 | -1.5 | 111 |
| 118 | ATR | 3.41 | 2.48 | 7.13 | 1.4 | -2.1 | -2.9 | 111 |
| 119 | AXL | 80.29 | 46.5 | 291.4 | 1.7 | -3.6 | -6.3 | 111 |
| 120 | BCR | 3.72 | 2.48 | 6.82 | 1.5 | -1.8 | -2.8 | 111 |
| 121 | BMP2K | 4.34 | 2.17 | 5.27 | 2. | -1.2 | -2.4 | 111 |
| 122 | BMPR1A | 7.44 | 4.34 | 8.37 | 1.7 | -1.1 | -1.9 | 111 |
| 123 | BMPR2 | 5.27 | 3.1 | 7.13 | 1.7 | -1.4 | -2.3 | 111 |

Continued on next page

**Table B.3 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 124 | BRAF | 7.13 | 3.41 | 11.16 | 2.1 | -1.6 | -3.3 | 111 |
| 125 | BRD2 | 8.06 | 4.03 | 10.23 | 2. | -1.3 | -2.5 | 111 |
| 126 | BUB1 | 5.58 | 6.82 | 8.37 | -1.2 | -1.5 | -1.2 | 111 |
| 127 | CAMK1 | 6.51 | 2.79 | 3.72 | 2.3 | 1.8 | -1.3 | 111 |
| 128 | CAMK2D | 7.75 | 4.34 | 8.68 | 1.8 | -1.1 | -2. | 111 |
| 129 | CASK | 3.1 | 0.93 | 2.17 | 3.3 | 1.4 | -2.3 | 111 |
| 130 | CDC2 | 37.51 | 41.54 | 29.14 | -1.1 | 1.3 | 1.4 | 111 |
| 131 | CDC2L2 | 5.89 | 3.72 | 8.37 | 1.6 | -1.4 | -2.2 | 111 |
| 132 | CDC2L5 | 6.51 | 3.1 | 10.54 | 2.1 | -1.6 | -3.4 | 111 |
| 133 | CDC42BPB | 3.72 | 1.86 | 6.2 | 2. | -1.7 | -3.3 | 111 |
| 134 | CDC7 | 4.03 | 1.86 | 5.89 | 2.2 | -1.5 | -3.2 | 111 |
| 135 | CDK10 | 5.27 | 3.41 | 11.16 | 1.5 | -2.1 | -3.3 | 111 |
| 136 | CDK2 | 7.13 | 5.58 | 8.06 | 1.3 | -1.1 | -1.4 | 111 |
| 137 | CDK8 | 6.51 | 3.72 | 11.47 | 1.8 | -1.8 | -3.1 | 111 |
| 138 | CHEK1 | 6.2 | 2.79 | 8.37 | 2.2 | -1.3 | -3. | 111 |
| 139 | CHUK | 16.74 | 8.68 | 21.39 | 1.9 | -1.3 | -2.5 | 111 |
| 140 | CLK1 | 4.96 | 1.55 | 9.3 | 3.2 | -1.9 | -6. | 111 |
| 141 | CLK2 | 5.27 | 2.48 | 6.82 | 2.1 | -1.3 | -2.8 | 111 |
| 142 | CLK4 | 9.3 | 3.41 | 13.02 | 2.7 | -1.4 | -3.8 | 111 |
| 143 | COL4A3BP | 9.61 | 4.34 | 8.06 | 2.2 | 1.2 | -1.9 | 111 |
| 144 | CPNE3 | 6.51 | 4.96 | 8.68 | 1.3 | -1.3 | -1.8 | 111 |
| 145 | CRKRS | 3.41 | 2.17 | 5.27 | 1.6 | -1.5 | -2.4 | 111 |
| 146 | CSK | 4.03 | 2.79 | 5.58 | 1.4 | -1.4 | -2. | 111 |
| 147 | CSNK1A1 | 18.91 | 8.37 | 29.45 | 2.3 | -1.6 | -3.5 | 111 |
| 148 | CSNK1D | 62.31 | 29.45 | 75.95 | 2.1 | -1.2 | -2.6 | 111 |
| 149 | CSNK1E | 16.74 | 11.78 | 26.97 | 1.4 | -1.6 | -2.3 | 111 |
| 150 | CSNK1G1 | 5.27 | 1.55 | 6.51 | 3.4 | -1.2 | -4.2 | 111 |
| 151 | CSNK1G3 | 11.47 | 6.51 | 15.81 | 1.8 | -1.4 | -2.4 | 111 |
| 152 | CSNK2A1 | 13.64 | 6.82 | 17.67 | 2. | -1.3 | -2.6 | 111 |
| 153 | CSNK2A2 | 4.34 | 0.62 | 6.2 | 7. | -1.4 | -10. | 111 |
| 154 | DAPK3 | 5.58 | 3.41 | 8.06 | 1.6 | -1.4 | -2.4 | 111 |
| 155 | DDR2 | 4.34 | 0. | 2.17 | 1400. | 2. | -700. | 111 |
| 156 | DYRK1A | 8.37 | 4.03 | 10.54 | 2.1 | -1.3 | -2.6 | 111 |
| 157 | DYRK2 | 2.79 | 1.86 | 4.03 | 1.5 | -1.4 | -2.2 | 111 |
| 158 | EGFR | 9.92 | 5.27 | 51.77 | 1.9 | -5.2 | -9.8 | 111 |
| 159 | EIF2AK1 | 7.75 | 6.51 | 8.99 | 1.2 | -1.2 | -1.4 | 111 |
| 160 | EIF2AK2 | 8.06 | 5.58 | 10.54 | 1.4 | -1.3 | -1.9 | 111 |
| 161 | EIF2AK3 | 2.48 | 0.62 | 5.27 | 4. | -2.1 | -8.5 | 111 |
| 162 | EIF2AK4 | 6.82 | 5.89 | 11.16 | 1.2 | -1.6 | -1.9 | 111 |
| 163 | EPHA2 | 6.82 | 4.65 | 27.9 | 1.5 | -4.1 | -6. | 111 |
| 164 | ERBB2 | 4.03 | 2.79 | 7.13 | 1.4 | -1.8 | -2.6 | 111 |
| 165 | ERN1 | 12.4 | 1.55 | 42.78 | 8. | -3.4 | -27.6 | 111 |

Continued on next page

**Table B.3 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 166 | FER | 8.06 | 4.65 | 11.78 | 1.7 | -1.5 | -2.5 | 111 |
| 167 | FGFR1 | 4.96 | 3.1 | 1.24 | 1.6 | 4. | 2.5 | 111 |
| 168 | FLJ13149 | 4.65 | 3.41 | 5.58 | 1.4 | -1.2 | -1.6 | 111 |
| 169 | FLJ21901 | 4.65 | 3.72 | 7.75 | 1.3 | -1.7 | -2.1 | 111 |
| 170 | FLJ23356 | 10.85 | 4.96 | 18.91 | 2.2 | -1.7 | -3.8 | 111 |
| 171 | GAK | 3.41 | 1.24 | 4.96 | 2.8 | -1.5 | -4. | 111 |
| 172 | GRK6 | 25.42 | 13.02 | 30.38 | 2. | -1.2 | -2.3 | 111 |
| 173 | GSG2 | 4.96 | 3.1 | 7.13 | 1.6 | -1.4 | -2.3 | 111 |
| 174 | GSK3A | 9.3 | 5.27 | 10.85 | 1.8 | -1.2 | -2.1 | 111 |
| 175 | GSK3B | 21.7 | 12.71 | 22.94 | 1.7 | -1.1 | -1.8 | 111 |
| 176 | HIPK3 | 5.89 | 2.79 | 7.44 | 2.1 | -1.3 | -2.7 | 111 |
| 177 | HUS1 | 11.47 | 5.27 | 19.53 | 2.2 | -1.7 | -3.7 | 111 |
| 178 | IGF1R | 3.41 | 1.55 | 7.75 | 2.2 | -2.3 | -5. | 111 |
| 179 | INSR | 1.86 | 0.31 | 4.65 | 6. | -2.5 | -15. | 111 |
| 180 | IRAK1 | 17.98 | 10.85 | 27.59 | 1.7 | -1.5 | -2.5 | 111 |
| 181 | JAK1 | 3.1 | 1.55 | 4.03 | 2. | -1.3 | -2.6 | 111 |
| 182 | KIAA0971 | 2.79 | 1.24 | 5.27 | 2.2 | -1.9 | -4.2 | 111 |
| 183 | LATS1 | 7.75 | 3.72 | 13.02 | 2.1 | -1.7 | -3.5 | 111 |
| 184 | LATS2 | 8.37 | 3.72 | 11.16 | 2.2 | -1.3 | -3. | 111 |
| 185 | LIMK2 | 5.89 | 3.41 | 7.44 | 1.7 | -1.3 | -2.2 | 111 |
| 186 | LMTK2 | 7.75 | 2.48 | 16.43 | 3.1 | -2.1 | -6.6 | 111 |
| 187 | MAP2K1 | 17.05 | 9.3 | 22.01 | 1.8 | -1.3 | -2.4 | 111 |
| 188 | MAP2K2 | 23.87 | 16.43 | 31.62 | 1.5 | -1.3 | -1.9 | 111 |
| 189 | MAP2K4 | 4.96 | 2.17 | 6.82 | 2.3 | -1.4 | -3.1 | 111 |
| 190 | MAP3K2 | 5.58 | 2.48 | 6.51 | 2.2 | -1.2 | -2.6 | 111 |
| 191 | MAP3K7 | 9.92 | 6.51 | 14.26 | 1.5 | -1.4 | -2.2 | 111 |
| 192 | MAP4K3 | 4.34 | 2.79 | 6.51 | 1.6 | -1.5 | -2.3 | 111 |
| 193 | MAP4K4 | 14.57 | 8.99 | 20.46 | 1.6 | -1.4 | -2.3 | 111 |
| 194 | MAP4K5 | 8.06 | 4.65 | 13.64 | 1.7 | -1.7 | -2.9 | 111 |
| 195 | MAPK1 | 15.81 | 8.68 | 14.88 | 1.8 | 1.1 | -1.7 | 111 |
| 196 | MAPK6 | 21.39 | 7.13 | 26.97 | 3. | -1.3 | -3.8 | 111 |
| 197 | MAPK8 | 7.44 | 3.41 | 11.47 | 2.2 | -1.5 | -3.4 | 111 |
| 198 | MAPK9 | 8.68 | 4.34 | 10.85 | 2. | -1.2 | -2.5 | 111 |
| 199 | MAPKAPK2 | 4.65 | 3.1 | 6.82 | 1.5 | -1.5 | -2.2 | 111 |
| 200 | MAPKAPK5 | 6.82 | 3.1 | 9.92 | 2.2 | -1.5 | -3.2 | 111 |
| 201 | MARK2 | 11.16 | 6.2 | 16.12 | 1.8 | -1.4 | -2.6 | 111 |
| 202 | MARK3 | 1.86 | 0.62 | 2.79 | 3. | -1.5 | -4.5 | 111 |
| 203 | MARK4 | 3.72 | 1.55 | 4.96 | 2.4 | -1.3 | -3.2 | 111 |
| 204 | MASTL | 13.64 | 5.89 | 17.05 | 2.3 | -1.2 | -2.9 | 111 |
| 205 | MELK | 35.96 | 21.7 | 39.06 | 1.7 | -1.1 | -1.8 | 111 |
| 206 | MET | 10.54 | 7.44 | 52.7 | 1.4 | -5. | -7.1 | 111 |
| 207 | MGC16169 | 3.41 | 1.55 | 4.96 | 2.2 | -1.5 | -3.2 | 111 |

Continued on next page

**Table B.3 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 208 | MKNK1 | 4.65 | 2.79 | 7.13 | 1.7 | -1.5 | -2.6 | 111 |
| 209 | MST1R | 5.27 | 3.72 | 17.67 | 1.4 | -3.4 | -4.8 | 111 |
| 210 | MST4 | 4.34 | 2.79 | 8.06 | 1.6 | -1.9 | -2.9 | 111 |
| 211 | MTOR | 1.86 | 0.62 | 2.79 | 3. | -1.5 | -4.5 | 111 |
| 212 | NEK1 | 2.17 | 0.31 | 3.1 | 7. | -1.4 | -10. | 111 |
| 213 | NEK3 | 3.41 | 0.62 | 4.96 | 5.5 | -1.5 | -8. | 111 |
| 214 | NEK7 | 9.92 | 4.65 | 11.16 | 2.1 | -1.1 | -2.4 | 111 |
| 215 | OXSR1 | 5.58 | 3.1 | 9.3 | 1.8 | -1.7 | -3. | 111 |
| 216 | PAK1 | 4.96 | 3.1 | 7.44 | 1.6 | -1.5 | -2.4 | 111 |
| 217 | PAK2 | 9.3 | 3.72 | 11.78 | 2.5 | -1.3 | -3.2 | 111 |
| 218 | PAN3 | 3.72 | 2.48 | 6.2 | 1.5 | -1.7 | -2.5 | 111 |
| 219 | PCTK1 | 11.78 | 6.2 | 13.64 | 1.9 | -1.2 | -2.2 | 111 |
| 220 | PCTK2 | 5.89 | 3.1 | 13.95 | 1.9 | -2.4 | -4.5 | 111 |
| 221 | PDIK1L | 3.72 | 0.93 | 5.58 | 4. | -1.5 | -6. | 111 |
| 222 | PDK1 | 3.72 | 2.48 | 5.89 | 1.5 | -1.6 | -2.4 | 111 |
| 223 | PFTK1 | 6.51 | 3.41 | 4.34 | 1.9 | 1.5 | -1.3 | 111 |
| 224 | PIK3R4 | 4.03 | 2.48 | 7.44 | 1.6 | -1.8 | -3. | 111 |
| 225 | PIM1 | 3.41 | 0.93 | 7.13 | 3.7 | -2.1 | -7.7 | 111 |
| 226 | PIM3 | 3.41 | 1.55 | 2.48 | 2.2 | 1.4 | -1.6 | 111 |
| 227 | PINK1 | 3.41 | 0.93 | 5.58 | 3.7 | -1.6 | -6. | 111 |
| 228 | PKMYT1 | 10.23 | 2.48 | 11.47 | 4.1 | -1.1 | -4.6 | 111 |
| 229 | PKN1 | 11.78 | 4.34 | 15.5 | 2.7 | -1.3 | -3.6 | 111 |
| 230 | PKN2 | 6.51 | 3.41 | 8.37 | 1.9 | -1.3 | -2.5 | 111 |
| 231 | PKN3 | 4.03 | 2.48 | 5.58 | 1.6 | -1.4 | -2.2 | 111 |
| 232 | PLK1 | 4.34 | 10.54 | 6.51 | -2.4 | -1.5 | 1.6 | 111 |
| 233 | PLK2 | 6.51 | 10.23 | 11.16 | -1.6 | -1.7 | -1.1 | 111 |
| 234 | PLK3 | 1.55 | 0.31 | 3.41 | 5. | -2.2 | -11. | 111 |
| 235 | PRKAA1 | 8.99 | 7.13 | 16.12 | 1.3 | -1.8 | -2.3 | 111 |
| 236 | PRKCI | 9.3 | 6.51 | 15.5 | 1.4 | -1.7 | -2.4 | 111 |
| 237 | PRKD3 | 5.58 | 3.72 | 6.82 | 1.5 | -1.2 | -1.8 | 111 |
| 238 | PRKDC | 36.58 | 19.84 | 64.79 | 1.8 | -1.8 | -3.3 | 111 |
| 239 | PRPF4B | 14.57 | 12.09 | 22.32 | 1.2 | -1.5 | -1.8 | 111 |
| 240 | PTK2 | 19.22 | 11.16 | 18.29 | 1.7 | 1.1 | -1.6 | 111 |
| 241 | PTK7 | 4.03 | 2.79 | 9.3 | 1.4 | -2.3 | -3.3 | 111 |
| 242 | RAF1 | 13.64 | 9.3 | 15.81 | 1.5 | -1.2 | -1.7 | 111 |
| 243 | RIOK2 | 7.44 | 4.34 | 9.92 | 1.7 | -1.3 | -2.3 | 111 |
| 244 | RIOK3 | 9.92 | 4.03 | 15.19 | 2.5 | -1.5 | -3.8 | 111 |
| 245 | RIPK1 | 4.65 | 3.41 | 7.13 | 1.4 | -1.5 | -2.1 | 111 |
| 246 | RIPK2 | 6.82 | 3.1 | 9.61 | 2.2 | -1.4 | -3.1 | 111 |
| 247 | RIPK4 | 3.72 | 4.96 | 6.2 | -1.3 | -1.7 | -1.2 | 111 |
| 248 | ROCK1 | 13.95 | 7.75 | 18.29 | 1.8 | -1.3 | -2.4 | 111 |
| 249 | ROCK2 | 8.06 | 5.27 | 10.85 | 1.5 | -1.3 | -2.1 | 111 |

Continued on next page

**Table B.3 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 250 | RPS6KA4 | 3.72 | 2.48 | 5.58 | 1.5 | -1.5 | -2.2 | 111 |
| 251 | RPS6KB1 | 4.65 | 2.79 | 6.82 | 1.7 | -1.5 | -2.4 | 111 |
| 252 | RPS6KB2 | 2.48 | 1.24 | 3.72 | 2. | -1.5 | -3. | 111 |
| 253 | RPS6KC1 | 4.96 | 2.79 | 7.13 | 1.8 | -1.4 | -2.6 | 111 |
| 254 | RYK | 4.03 | 2.79 | 6.2 | 1.4 | -1.5 | -2.2 | 111 |
| 255 | SCYL1 | 5.27 | 2.48 | 6.82 | 2.1 | -1.3 | -2.8 | 111 |
| 256 | SCYL2 | 10.54 | 4.65 | 14.88 | 2.3 | -1.4 | -3.2 | 111 |
| 257 | SGK | 0.62 | 3.1 | 1.55 | -5. | -2.5 | 2. | 111 |
| 258 | SLK | 8.06 | 5.27 | 12.09 | 1.5 | -1.5 | -2.3 | 111 |
| 259 | SMG1 | 19.53 | 5.58 | 32.86 | 3.5 | -1.7 | -5.9 | 111 |
| 260 | SNF1LK2 | 6.82 | 2.79 | 7.75 | 2.4 | -1.1 | -2.8 | 111 |
| 261 | SNRK | 4.34 | 1.24 | 5.27 | 3.5 | -1.2 | -4.2 | 111 |
| 262 | SRC | 2.17 | 1.24 | 4.96 | 1.8 | -2.3 | -4. | 111 |
| 263 | SRPK1 | 8.68 | 7.44 | 11.47 | 1.2 | -1.3 | -1.5 | 111 |
| 264 | SRPK2 | 8.06 | 4.34 | 9.92 | 1.9 | -1.2 | -2.3 | 111 |
| 265 | STK11 | 2.48 | 1.24 | 3.41 | 2. | -1.4 | -2.8 | 111 |
| 266 | STK24 | 15.81 | 8.99 | 27.9 | 1.8 | -1.8 | -3.1 | 111 |
| 267 | STK39 | 9.92 | 4.03 | 13.95 | 2.5 | -1.4 | -3.5 | 111 |
| 268 | STK40 | 2.79 | 1.24 | 4.34 | 2.2 | -1.6 | -3.5 | 111 |
| 269 | TAF1 | 2.48 | 1.24 | 3.72 | 2. | -1.5 | -3. | 111 |
| 270 | TAF1L | 3.1 | 1.24 | 4.34 | 2.5 | -1.4 | -3.5 | 111 |
| 271 | TAOK1 | 10.23 | 5.58 | 12.4 | 1.8 | -1.2 | -2.2 | 111 |
| 272 | TAOK3 | 4.03 | 2.17 | 9.3 | 1.9 | -2.3 | -4.3 | 111 |
| 273 | TGFBR2 | 25.73 | 53.63 | 33.79 | -2.1 | -1.3 | 1.6 | 111 |
| 274 | TLK1 | 4.34 | 2.79 | 5.58 | 1.6 | -1.3 | -2. | 111 |
| 275 | TLK2 | 5.58 | 2.79 | 7.44 | 2. | -1.3 | -2.7 | 111 |
| 276 | TP53RK | 17.36 | 8.68 | 20.77 | 2. | -1.2 | -2.4 | 111 |
| 277 | TRIB1 | 1.55 | 0.62 | 2.48 | 2.5 | -1.6 | -4. | 111 |
| 278 | TRIB2 | 1.55 | 6.82 | 0.31 | -4.4 | 5. | 22. | 111 |
| 279 | TRIB3 | 57.04 | 5.58 | 68.2 | 10.2 | -1.2 | -12.2 | 111 |
| 280 | TRIO | 10.85 | 5.89 | 19.53 | 1.8 | -1.8 | -3.3 | 111 |
| 281 | TRPM7 | 5.58 | 1.86 | 8.06 | 3. | -1.4 | -4.3 | 111 |
| 282 | TTBK2 | 4.03 | 0.62 | 6.51 | 6.5 | -1.6 | -10.5 | 111 |
| 283 | TTK | 11.16 | 13.33 | 14.88 | -1.2 | -1.3 | -1.1 | 111 |
| 284 | UHMK1 | 12.09 | 7.13 | 17.67 | 1.7 | -1.5 | -2.5 | 111 |
| 285 | ULK3 | 6.82 | 3.41 | 9.3 | 2. | -1.4 | -2.7 | 111 |
| 286 | VRK1 | 10.23 | 6.2 | 13.64 | 1.6 | -1.3 | -2.2 | 111 |
| 287 | VRK2 | 3.72 | 1.55 | 5.58 | 2.4 | -1.5 | -3.6 | 111 |
| 288 | VRK3 | 4.34 | 1.86 | 5.58 | 2.3 | -1.3 | -3. | 111 |
| 289 | WNK1 | 20.15 | 8.37 | 21.7 | 2.4 | -1.1 | -2.6 | 111 |
| 290 | YES1 | 24.18 | 14.88 | 31.62 | 1.6 | -1.3 | -2.1 | 111 |

Table B.4: List of significantly changed gene in the group: (1 - $H827Par$, 2 - $H827ER20$, 3 - $HR827ER40$), along with pairwise fold changes and significant change markers

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 1 | NEK2 | 1.86 | 2.17 | 1.24 | -1.2 | 1.5 | 1.8 | 001 |
| 2 | NUAK2 | 2.17 | 2.48 | 1.55 | -1.1 | 1.4 | 1.6 | 001 |
| 3 | BUB1B | 3.72 | 3.1 | 2.48 | 1.2 | 1.5 | 1.2 | 010 |
| 4 | CSK | 0.93 | 1.55 | 1.86 | -1.7 | -2. | -1.2 | 010 |
| 5 | AURKB | 8.99 | 8.68 | 6.2 | 1. | 1.4 | 1.4 | 011 |
| 6 | BUB1 | 4.34 | 4.03 | 2.79 | 1.1 | 1.6 | 1.4 | 011 |
| 7 | TTK | 4.03 | 4.65 | 3.1 | -1.2 | 1.3 | 1.5 | 011 |
| 8 | ATR | 0.93 | 1.86 | 1.55 | -2. | -1.7 | 1.2 | 100 |
| 9 | BMP2K | 1.24 | 2.17 | 1.55 | -1.8 | -1.2 | 1.4 | 100 |
| 10 | BMPR1A | 1.24 | 2.17 | 1.86 | -1.8 | -1.5 | 1.2 | 100 |
| 11 | BMPR1B | 0.31 | 1.24 | 0.62 | -4. | -2. | 2. | 100 |
| 12 | BMPR2 | 1.86 | 3.1 | 2.48 | -1.7 | -1.3 | 1.2 | 100 |
| 13 | CDK9 | 1.86 | 3.1 | 2.48 | -1.7 | -1.3 | 1.2 | 100 |
| 14 | CPNE3 | 2.17 | 3.1 | 2.48 | -1.4 | -1.1 | 1.2 | 100 |
| 15 | CSNK1G1 | 0.62 | 1.55 | 1.24 | -2.5 | -2. | 1.2 | 100 |
| 16 | DYRK1A | 0.93 | 1.86 | 1.24 | -2. | -1.3 | 1.5 | 100 |
| 17 | FER | 0.31 | 1.24 | 0.93 | -4. | -3. | 1.3 | 100 |
| 18 | ILK | 0.93 | 1.86 | 1.24 | -2. | -1.3 | 1.5 | 100 |
| 19 | LOC91461 | 0. | 0.93 | 0.62 | -300. | -200. | 1.5 | 100 |
| 20 | MAP3K1 | 2.48 | 3.41 | 2.79 | -1.4 | -1.1 | 1.2 | 100 |
| 21 | MAP3K14 | 0.31 | 1.24 | 0.93 | -4. | -3. | 1.3 | 100 |
| 22 | MAPKAPK3 | 1.55 | 2.79 | 2.17 | -1.8 | -1.4 | 1.3 | 100 |
| 23 | MKNK1 | 2.17 | 3.41 | 2.79 | -1.6 | -1.3 | 1.2 | 100 |
| 24 | MST4 | 5.27 | 6.51 | 5.89 | -1.2 | -1.1 | 1.1 | 100 |
| 25 | PAN3 | 0.93 | 1.86 | 1.55 | -2. | -1.7 | 1.2 | 100 |
| 26 | PRKCD | 3.1 | 4.34 | 3.72 | -1.4 | -1.2 | 1.2 | 100 |
| 27 | PRKCE | 0.93 | 2.17 | 1.55 | -2.3 | -1.7 | 1.4 | 100 |
| 28 | PRKCZ | 0.93 | 1.86 | 1.24 | -2. | -1.3 | 1.5 | 100 |
| 29 | RIPK4 | 2.17 | 3.1 | 2.79 | -1.4 | -1.3 | 1.1 | 100 |
| 30 | RPS6KA1 | 2.17 | 3.1 | 2.79 | -1.4 | -1.3 | 1.1 | 100 |
| 31 | STK17B | 3.1 | 4.34 | 3.72 | -1.4 | -1.2 | 1.2 | 100 |
| 32 | STYK1 | 0.62 | 1.55 | 0.93 | -2.5 | -1.5 | 1.7 | 100 |
| 33 | TAF1L | 0.93 | 1.86 | 1.55 | -2. | -1.7 | 1.2 | 100 |
| 34 | TLK1 | 1.24 | 2.17 | 1.55 | -1.8 | -1.2 | 1.4 | 100 |
| 35 | TRIB2 | 0. | 0.93 | 0.62 | -300. | -200. | 1.5 | 100 |
| 36 | CDC2 | 16.12 | 17.98 | 16.74 | -1.1 | -1. | 1.1 | 101 |
| 37 | DAPK1 | 2.79 | 4.34 | 3.41 | -1.6 | -1.2 | 1.3 | 101 |
| 38 | EIF2AK4 | 3.1 | 4.03 | 2.79 | -1.3 | 1.1 | 1.4 | 101 |
| 39 | FYN | 1.55 | 3.41 | 2.17 | -2.2 | -1.4 | 1.6 | 101 |

Continued on next page

**Table B.4 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 40 | MAPK9 | 2.17 | 3.72 | 2.79 | -1.7 | -1.3 | 1.3 | 101 |
| 41 | MELK | 3.41 | 4.34 | 3.1 | -1.3 | 1.1 | 1.4 | 101 |
| 42 | PCTK2 | 1.24 | 2.79 | 1.86 | -2.2 | -1.5 | 1.5 | 101 |
| 43 | PDIK1L | 1.24 | 2.48 | 1.55 | -2. | -1.2 | 1.6 | 101 |
| 44 | PRKACB | 2.17 | 3.41 | 2.48 | -1.6 | -1.1 | 1.4 | 101 |
| 45 | PRKCA | 0.93 | 2.48 | 1.55 | -2.7 | -1.7 | 1.6 | 101 |
| 46 | ROCK2 | 2.79 | 4.34 | 3.41 | -1.6 | -1.2 | 1.3 | 101 |
| 47 | STK25 | 1.86 | 3.41 | 2.48 | -1.8 | -1.3 | 1.4 | 101 |
| 48 | TBRG4 | 1.55 | 3.1 | 2.17 | -2. | -1.4 | 1.4 | 101 |
| 49 | AAK1 | 2.17 | 3.41 | 3.41 | -1.6 | -1.6 | -1. | 110 |
| 50 | ABL2 | 0.93 | 2.17 | 1.86 | -2.3 | -2. | 1.2 | 110 |
| 51 | ACVR1 | 0.93 | 2.48 | 2.17 | -2.7 | -2.3 | 1.1 | 110 |
| 52 | ACVR1B | 0.93 | 2.17 | 2.17 | -2.3 | -2.3 | -1. | 110 |
| 53 | ADCK2 | 0.93 | 1.86 | 1.86 | -2. | -2. | -1. | 110 |
| 54 | ALPK1 | 1.55 | 3.72 | 4.03 | -2.4 | -2.6 | -1.1 | 110 |
| 55 | ARAF | 1.86 | 3.72 | 3.1 | -2. | -1.7 | 1.2 | 110 |
| 56 | BCKDK | 2.17 | 3.72 | 3.1 | -1.7 | -1.4 | 1.2 | 110 |
| 57 | BCR | 2.17 | 3.72 | 3.1 | -1.7 | -1.4 | 1.2 | 110 |
| 58 | BRDT | 1.24 | 2.79 | 2.17 | -2.2 | -1.8 | 1.3 | 110 |
| 59 | CAMKK2 | 1.55 | 3.1 | 3.1 | -2. | -2. | -1. | 110 |
| 60 | CASK | 1.24 | 2.48 | 2.17 | -2. | -1.8 | 1.1 | 110 |
| 61 | CDC2L5 | 1.55 | 2.79 | 2.79 | -1.8 | -1.8 | -1. | 110 |
| 62 | CDC42BPB | 1.86 | 4.34 | 3.72 | -2.3 | -2. | 1.2 | 110 |
| 63 | CDK6 | 2.17 | 4.96 | 4.65 | -2.3 | -2.1 | 1.1 | 110 |
| 64 | CDK7 | 1.55 | 3.72 | 3.1 | -2.4 | -2. | 1.2 | 110 |
| 65 | CDK8 | 1.86 | 3.72 | 3.1 | -2. | -1.7 | 1.2 | 110 |
| 66 | CHUK | 2.48 | 4.65 | 4.34 | -1.9 | -1.8 | 1.1 | 110 |
| 67 | CLK2 | 1.55 | 3.1 | 2.48 | -2. | -1.6 | 1.2 | 110 |
| 68 | COL4A3BP | 3.41 | 7.13 | 6.51 | -2.1 | -1.9 | 1.1 | 110 |
| 69 | CRKRS | 0.93 | 2.17 | 1.86 | -2.3 | -2. | 1.2 | 110 |
| 70 | CSNK1A1 | 3.1 | 6.82 | 6.2 | -2.2 | -2. | 1.1 | 110 |
| 71 | CSNK1G3 | 2.48 | 5.89 | 5.27 | -2.4 | -2.1 | 1.1 | 110 |
| 72 | CSNK2A2 | 1.24 | 3.1 | 2.79 | -2.5 | -2.2 | 1.1 | 110 |
| 73 | DAPK3 | 1.86 | 4.34 | 3.72 | -2.3 | -2. | 1.2 | 110 |
| 74 | DDR1 | 1.86 | 4.03 | 3.41 | -2.2 | -1.8 | 1.2 | 110 |
| 75 | EIF2AK1 | 5.89 | 10.54 | 10.23 | -1.8 | -1.7 | 1. | 110 |
| 76 | EIF2AK2 | 5.27 | 9.3 | 8.99 | -1.8 | -1.7 | 1. | 110 |
| 77 | EIF2AK3 | 0.62 | 2.48 | 1.86 | -4. | -3. | 1.3 | 110 |
| 78 | ERN1 | 0.93 | 4.65 | 4.03 | -5. | -4.3 | 1.2 | 110 |
| 79 | FLJ21901 | 1.55 | 3.1 | 2.79 | -2. | -1.8 | 1.1 | 110 |
| 80 | GAK | 1.55 | 2.79 | 2.79 | -1.8 | -1.8 | -1. | 110 |
| 81 | GRK6 | 3.41 | 6.82 | 6.82 | -2. | -2. | -1. | 110 |

Continued on next page

**Table B.4 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 82 | GSK3A | 1.24 | 2.79 | 2.48 | -2.2 | -2. | 1.1 | 110 |
| 83 | HIPK3 | 1.86 | 3.72 | 3.1 | -2. | -1.7 | 1.2 | 110 |
| 84 | INSR | 0.93 | 3.41 | 2.79 | -3.7 | -3. | 1.2 | 110 |
| 85 | IRAK2 | 0.93 | 4.34 | 3.72 | -4.7 | -4. | 1.2 | 110 |
| 86 | LMTK2 | 1.55 | 4.03 | 4.03 | -2.6 | -2.6 | -1. | 110 |
| 87 | LYN | 0.93 | 3.1 | 3.1 | -3.3 | -3.3 | -1. | 110 |
| 88 | MAP2K1 | 8.68 | 17.36 | 16.74 | -2. | -1.9 | 1. | 110 |
| 89 | MAP2K3 | 2.17 | 5.89 | 5.27 | -2.7 | -2.4 | 1.1 | 110 |
| 90 | MAP2K4 | 1.86 | 3.1 | 2.79 | -1.7 | -1.5 | 1.1 | 110 |
| 91 | MAP2K7 | 1.55 | 3.72 | 3.1 | -2.4 | -2. | 1.2 | 110 |
| 92 | MAP3K2 | 2.79 | 4.65 | 4.34 | -1.7 | -1.6 | 1.1 | 110 |
| 93 | MAP3K5 | 0.93 | 3.1 | 3.1 | -3.3 | -3.3 | -1. | 110 |
| 94 | MAP3K7 | 2.48 | 4.03 | 3.72 | -1.6 | -1.5 | 1.1 | 110 |
| 95 | MAPK6 | 4.34 | 10.23 | 9.92 | -2.4 | -2.3 | 1. | 110 |
| 96 | MAPK8 | 1.55 | 3.72 | 3.1 | -2.4 | -2. | 1.2 | 110 |
| 97 | MAPKAPK2 | 2.48 | 4.03 | 4.03 | -1.6 | -1.6 | -1. | 110 |
| 98 | MAPKAPK5 | 2.79 | 5.58 | 4.96 | -2. | -1.8 | 1.1 | 110 |
| 99 | MARK1 | 0.31 | 1.24 | 1.55 | -4. | -5. | -1.2 | 110 |
| 100 | MARK4 | 1.24 | 2.48 | 2.17 | -2. | -1.8 | 1.1 | 110 |
| 101 | MAST2 | 1.55 | 3.1 | 2.48 | -2. | -1.6 | 1.2 | 110 |
| 102 | MGC16169 | 1.55 | 3.41 | 2.79 | -2.2 | -1.8 | 1.2 | 110 |
| 103 | MINK1 | 3.1 | 5.27 | 4.96 | -1.7 | -1.6 | 1.1 | 110 |
| 104 | MLKL | 1.24 | 3.1 | 2.79 | -2.5 | -2.2 | 1.1 | 110 |
| 105 | MST1R | 1.24 | 2.79 | 2.48 | -2.2 | -2. | 1.1 | 110 |
| 106 | NEK7 | 4.34 | 8.37 | 7.75 | -1.9 | -1.8 | 1.1 | 110 |
| 107 | OXSR1 | 1.55 | 3.1 | 2.48 | -2. | -1.6 | 1.2 | 110 |
| 108 | PAK2 | 3.1 | 4.65 | 4.34 | -1.5 | -1.4 | 1.1 | 110 |
| 109 | PAK4 | 1.86 | 3.41 | 3.1 | -1.8 | -1.7 | 1.1 | 110 |
| 110 | PDK1 | 2.48 | 4.96 | 4.34 | -2. | -1.8 | 1.1 | 110 |
| 111 | PDK4 | 0.31 | 1.24 | 1.24 | -4. | -4. | -1. | 110 |
| 112 | PIM3 | 1.55 | 6.2 | 5.89 | -4. | -3.8 | 1.1 | 110 |
| 113 | PKN1 | 3.1 | 8.06 | 7.75 | -2.6 | -2.5 | 1. | 110 |
| 114 | PRKCH | 1.24 | 2.79 | 2.79 | -2.2 | -2.2 | -1. | 110 |
| 115 | PRKD3 | 1.55 | 2.79 | 2.79 | -1.8 | -1.8 | -1. | 110 |
| 116 | PRKDC | 4.03 | 5.58 | 4.96 | -1.4 | -1.2 | 1.1 | 110 |
| 117 | PRPF4B | 2.48 | 4.96 | 4.65 | -2. | -1.9 | 1.1 | 110 |
| 118 | RAF1 | 2.17 | 4.34 | 4.03 | -2. | -1.9 | 1.1 | 110 |
| 119 | RIOK2 | 1.55 | 2.48 | 2.79 | -1.6 | -1.8 | -1.1 | 110 |
| 120 | RIOK3 | 1.86 | 5.58 | 4.96 | -3. | -2.7 | 1.1 | 110 |
| 121 | RIPK1 | 1.24 | 2.79 | 2.48 | -2.2 | -2. | 1.1 | 110 |
| 122 | RIPK2 | 8.06 | 23.56 | 22.94 | -2.9 | -2.8 | 1. | 110 |
| 123 | ROCK1 | 2.48 | 4.34 | 4.34 | -1.8 | -1.8 | -1. | 110 |

Continued on next page

**Table B.4 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 124 | RPS6KC1 | 1.55 | 3.41 | 2.79 | -2.2 | -1.8 | 1.2 | 110 |
| 125 | SCYL1 | 1.86 | 3.41 | 3.41 | -1.8 | -1.8 | -1. | 110 |
| 126 | SCYL2 | 2.79 | 5.58 | 5.27 | -2. | -1.9 | 1.1 | 110 |
| 127 | SCYL3 | 0.62 | 1.86 | 1.55 | -3. | -2.5 | 1.2 | 110 |
| 128 | SGK3 | 1.55 | 2.48 | 2.48 | -1.6 | -1.6 | -1. | 110 |
| 129 | SMG1 | 2.48 | 4.96 | 4.34 | -2. | -1.8 | 1.1 | 110 |
| 130 | SNF1LK2 | 1.55 | 2.79 | 2.48 | -1.8 | -1.6 | 1.1 | 110 |
| 131 | SRPK1 | 5.27 | 11.47 | 10.85 | -2.2 | -2.1 | 1.1 | 110 |
| 132 | SRPK2 | 3.72 | 5.89 | 5.89 | -1.6 | -1.6 | -1. | 110 |
| 133 | STK16 | 0.93 | 1.86 | 1.86 | -2. | -2. | -1. | 110 |
| 134 | STK32A | 0.62 | 1.86 | 1.86 | -3. | -3. | -1. | 110 |
| 135 | STK35 | 0.93 | 1.86 | 1.86 | -2. | -2. | -1. | 110 |
| 136 | STK38 | 2.79 | 4.03 | 3.72 | -1.4 | -1.3 | 1.1 | 110 |
| 137 | STK38L | 1.86 | 3.41 | 3.1 | -1.8 | -1.7 | 1.1 | 110 |
| 138 | STK39 | 6.82 | 15.5 | 14.88 | -2.3 | -2.2 | 1. | 110 |
| 139 | STK40 | 1.55 | 4.65 | 4.03 | -3. | -2.6 | 1.2 | 110 |
| 140 | TAOK1 | 2.17 | 4.03 | 3.41 | -1.9 | -1.6 | 1.2 | 110 |
| 141 | TAOK3 | 2.48 | 4.96 | 4.34 | -2. | -1.8 | 1.1 | 110 |
| 142 | TLK2 | 0.93 | 2.48 | 1.86 | -2.7 | -2. | 1.3 | 110 |
| 143 | TP53RK | 3.41 | 7.13 | 6.51 | -2.1 | -1.9 | 1.1 | 110 |
| 144 | TRIB1 | 1.86 | 5.58 | 5.58 | -3. | -3. | -1. | 110 |
| 145 | TRIB3 | 1.86 | 5.58 | 6.2 | -3. | -3.3 | -1.1 | 110 |
| 146 | UHMK1 | 4.96 | 8.68 | 8.06 | -1.8 | -1.6 | 1.1 | 110 |
| 147 | ULK1 | 0.93 | 2.17 | 1.86 | -2.3 | -2. | 1.2 | 110 |
| 148 | ULK3 | 1.55 | 3.41 | 2.79 | -2.2 | -1.8 | 1.2 | 110 |
| 149 | VRK2 | 1.86 | 4.03 | 3.72 | -2.2 | -2. | 1.1 | 110 |
| 150 | ADRBK1 | 3.72 | 9.92 | 8.68 | -2.7 | -2.3 | 1.1 | 111 |
| 151 | AKT2 | 3.72 | 5.89 | 4.96 | -1.6 | -1.3 | 1.2 | 111 |
| 152 | BRAF | 2.17 | 4.03 | 3.1 | -1.9 | -1.4 | 1.3 | 111 |
| 153 | BRD2 | 1.86 | 4.03 | 3.1 | -2.2 | -1.7 | 1.3 | 111 |
| 154 | CAMK2D | 4.65 | 10.23 | 8.99 | -2.2 | -1.9 | 1.1 | 111 |
| 155 | CAMK2G | 1.86 | 3.72 | 2.79 | -2. | -1.5 | 1.3 | 111 |
| 156 | CDK10 | 3.72 | 7.75 | 6.82 | -2.1 | -1.8 | 1.1 | 111 |
| 157 | CDK4 | 243.97 | 318.37 | 283.03 | -1.3 | -1.2 | 1.1 | 111 |
| 158 | CSNK1D | 15.81 | 35.03 | 34.1 | -2.2 | -2.2 | 1. | 111 |
| 159 | CSNK1E | 4.03 | 11.47 | 8.99 | -2.8 | -2.2 | 1.3 | 111 |
| 160 | CSNK2A1 | 2.48 | 6.82 | 5.58 | -2.8 | -2.2 | 1.2 | 111 |
| 161 | DYRK2 | 1.86 | 4.34 | 3.1 | -2.3 | -1.7 | 1.4 | 111 |
| 162 | EGFR | 78.43 | 110.98 | 88.66 | -1.4 | -1.1 | 1.3 | 111 |
| 163 | EPHA2 | 1.86 | 5.58 | 4.34 | -3. | -2.3 | 1.3 | 111 |
| 164 | ERBB3 | 6.51 | 10.85 | 9.61 | -1.7 | -1.5 | 1.1 | 111 |
| 165 | FLJ23356 | 1.55 | 3.72 | 2.79 | -2.4 | -1.8 | 1.3 | 111 |

Continued on next page

**Table B.4 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 166 | GSK3B | 6.2 | 12.4 | 10.23 | -2. | -1.6 | 1.2 | 111 |
| 167 | HSPB8 | 0.93 | 3.72 | 2.48 | -4. | -2.7 | 1.5 | 111 |
| 168 | HUS1 | 2.48 | 5.27 | 4.34 | -2.1 | -1.8 | 1.2 | 111 |
| 169 | IRAK1 | 9.3 | 12.09 | 10.23 | -1.3 | -1.1 | 1.2 | 111 |
| 170 | JAK1 | 4.03 | 8.99 | 8.06 | -2.2 | -2. | 1.1 | 111 |
| 171 | LATS1 | 1.55 | 4.65 | 3.72 | -3. | -2.4 | 1.3 | 111 |
| 172 | LIMK1 | 4.96 | 7.13 | 6.2 | -1.4 | -1.2 | 1.1 | 111 |
| 173 | MAP2K2 | 7.13 | 13.33 | 11.47 | -1.9 | -1.6 | 1.2 | 111 |
| 174 | MAP4K3 | 2.48 | 5.58 | 4.65 | -2.2 | -1.9 | 1.2 | 111 |
| 175 | MAP4K4 | 6.2 | 24.49 | 20.77 | -3.9 | -3.3 | 1.2 | 111 |
| 176 | MAP4K5 | 4.96 | 7.75 | 6.82 | -1.6 | -1.4 | 1.1 | 111 |
| 177 | MAPK1 | 4.03 | 6.2 | 5.27 | -1.5 | -1.3 | 1.2 | 111 |
| 178 | MAPK13 | 3.72 | 9.3 | 8.37 | -2.5 | -2.2 | 1.1 | 111 |
| 179 | MAPK14 | 3.1 | 6.51 | 5.58 | -2.1 | -1.8 | 1.2 | 111 |
| 180 | MAPK3 | 3.1 | 6.2 | 4.65 | -2. | -1.5 | 1.3 | 111 |
| 181 | MARK2 | 4.65 | 10.23 | 8.68 | -2.2 | -1.9 | 1.2 | 111 |
| 182 | MET | 40.61 | 499.41 | 426.25 | -12.3 | -10.5 | 1.2 | 111 |
| 183 | MKNK2 | 3.1 | 9.92 | 8.99 | -3.2 | -2.9 | 1.1 | 111 |
| 184 | MYLK | 2.48 | 1.24 | 0.31 | 2. | 8. | 4. | 111 |
| 185 | NRBP1 | 4.96 | 8.99 | 6.82 | -1.8 | -1.4 | 1.3 | 111 |
| 186 | PAK1 | 4.34 | 8.99 | 7.75 | -2.1 | -1.8 | 1.2 | 111 |
| 187 | PCTK1 | 3.41 | 5.58 | 4.65 | -1.6 | -1.4 | 1.2 | 111 |
| 188 | PIM1 | 3.41 | 6.82 | 5.27 | -2. | -1.5 | 1.3 | 111 |
| 189 | PKN2 | 4.34 | 7.44 | 6.2 | -1.7 | -1.4 | 1.2 | 111 |
| 190 | PLK2 | 4.34 | 8.68 | 6.51 | -2. | -1.5 | 1.3 | 111 |
| 191 | PRKAA1 | 5.58 | 11.47 | 9.3 | -2.1 | -1.7 | 1.2 | 111 |
| 192 | PRKAA2 | 2.17 | 4.34 | 3.41 | -2. | -1.6 | 1.3 | 111 |
| 193 | PRKACA | 5.58 | 8.68 | 7.75 | -1.6 | -1.4 | 1.1 | 111 |
| 194 | PRKCI | 5.58 | 9.61 | 7.75 | -1.7 | -1.4 | 1.2 | 111 |
| 195 | PTK2 | 5.89 | 9.61 | 8.06 | -1.6 | -1.4 | 1.2 | 111 |
| 196 | RPS6KA3 | 2.79 | 7.44 | 6.2 | -2.7 | -2.2 | 1.2 | 111 |
| 197 | SGK | 5.27 | 38.75 | 41.85 | -7.4 | -7.9 | -1.1 | 111 |
| 198 | SLK | 2.17 | 4.03 | 3.1 | -1.9 | -1.4 | 1.3 | 111 |
| 199 | SRC | 2.17 | 5.27 | 4.34 | -2.4 | -2. | 1.2 | 111 |
| 200 | STK17A | 4.03 | 8.06 | 5.27 | -2. | -1.3 | 1.5 | 111 |
| 201 | STK24 | 3.41 | 7.44 | 6.51 | -2.2 | -1.9 | 1.1 | 111 |
| 202 | TBK1 | 34.72 | 58.28 | 54.87 | -1.7 | -1.6 | 1.1 | 111 |
| 203 | TGFBR2 | 4.03 | 6.51 | 5.58 | -1.6 | -1.4 | 1.2 | 111 |
| 204 | TRIO | 7.75 | 12.71 | 10.54 | -1.6 | -1.4 | 1.2 | 111 |
| 205 | WNK1 | 5.89 | 11.78 | 10.85 | -2. | -1.8 | 1.1 | 111 |
| 206 | YES1 | 6.2 | 8.99 | 7.13 | -1.4 | -1.1 | 1.3 | 111 |

Table B.5: List of significantly changed gene in the group: (1 - $A549$, 2 - $H358$, 3 - $H2122$), along with pairwise fold changes and significant change markers

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 1 | ACVR1 | 0.93 | 0.31 | 1.24 | 3. | -1.3 | -4. | 001 |
| 2 | ACVR1B | 0.62 | 0. | 1.24 | 200. | -2. | -400. | 001 |
| 3 | CDC42BPA | 0.62 | 0. | 1.24 | 200. | -2. | -400. | 001 |
| 4 | CDK5 | 0.93 | 0.31 | 1.55 | 3. | -1.7 | -5. | 001 |
| 5 | EEF2K | 0.31 | 0. | 0.93 | 100. | -3. | -300. | 001 |
| 6 | PHKG2 | 0.93 | 0.62 | 1.55 | 1.5 | -1.7 | -2.5 | 001 |
| 7 | PKMYT1 | 1.24 | 0.93 | 1.86 | 1.3 | -1.5 | -2. | 001 |
| 8 | SGK269 | 0.62 | 0.31 | 1.24 | 2. | -2. | -4. | 001 |
| 9 | STK4 | 0.31 | 0. | 0.93 | 100. | -3. | -300. | 001 |
| 10 | ZAK | 0.31 | 0. | 0.93 | 100. | -3. | -300. | 001 |
| 11 | EPHA1 | 0. | 0.31 | 0.93 | -100. | -300. | -3. | 010 |
| 12 | MAP2K6 | 1.55 | 0.93 | 0.62 | 1.7 | 2.5 | 1.5 | 010 |
| 13 | MARK3 | 0.31 | 0.62 | 1.24 | -2. | -4. | -2. | 010 |
| 14 | PDK3 | 0.93 | 1.55 | 1.86 | -1.7 | -2. | -1.2 | 010 |
| 15 | PINK1 | 0.93 | 1.24 | 1.86 | -1.3 | -2. | -1.5 | 010 |
| 16 | PTK7 | 0. | 0.62 | 0.93 | -200. | -300. | -1.5 | 010 |
| 17 | RIOK1 | 0. | 0.62 | 0.93 | -200. | -300. | -1.5 | 010 |
| 18 | SCYL3 | 0.31 | 0.93 | 1.55 | -3. | -5. | -1.7 | 010 |
| 19 | AAK1 | 1.55 | 1.55 | 3.72 | -1. | -2.4 | -2.4 | 011 |
| 20 | ACVR2A | 0.31 | 0.31 | 1.24 | -1. | -4. | -4. | 011 |
| 21 | ACVR2B | 0. | 0. | 0.93 | -1. | -300. | -300. | 011 |
| 22 | ALPK1 | 0. | 0.31 | 2.17 | -100. | -700. | -7. | 011 |
| 23 | ATR | 0.93 | 0.93 | 3.72 | -1. | -4. | -4. | 011 |
| 24 | BCKDK | 2.48 | 1.86 | 6.51 | 1.3 | -2.6 | -3.5 | 011 |
| 25 | BRAF | 0.93 | 0.93 | 3.72 | -1. | -4. | -4. | 011 |
| 26 | BUB1 | 2.79 | 2.48 | 3.72 | 1.1 | -1.3 | -1.5 | 011 |
| 27 | CABC1 | 0. | 0. | 1.55 | -1. | -500. | -500. | 011 |
| 28 | CAMK1 | 0.31 | 0. | 2.79 | 100. | -9. | -900. | 011 |
| 29 | CAMK2G | 1.55 | 0.93 | 10.85 | 1.7 | -7. | -11.7 | 011 |
| 30 | CAMKK2 | 1.55 | 0.93 | 6.2 | 1.7 | -4. | -6.7 | 011 |
| 31 | CASK | 0.62 | 0.93 | 2.17 | -1.5 | -3.5 | -2.3 | 011 |
| 32 | CDC2L2 | 0.93 | 1.24 | 2.79 | -1.3 | -3. | -2.2 | 011 |
| 33 | CDC2L5 | 1.86 | 1.24 | 7.13 | 1.5 | -3.8 | -5.8 | 011 |
| 34 | CDC2L6 | 1.24 | 0.62 | 2.48 | 2. | -2. | -4. | 011 |
| 35 | CDK10 | 2.17 | 1.86 | 7.44 | 1.2 | -3.4 | -4. | 011 |
| 36 | CDK2 | 1.55 | 1.55 | 2.48 | -1. | -1.6 | -1.6 | 011 |
| 37 | CDK7 | 0.93 | 0.93 | 7.44 | -1. | -8. | -8. | 011 |
| 38 | CDK8 | 1.55 | 0.93 | 6.2 | 1.7 | -4. | -6.7 | 011 |
| 39 | CDK9 | 1.24 | 1.55 | 4.34 | -1.2 | -3.5 | -2.8 | 011 |
| 40 | CDKL1 | 0.31 | 0. | 5.89 | 100. | -19. | -1900. | 011 |

Continued on next page

**Table B.5 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 41 | CHEK2 | 0.93 | 0.93 | 3.72 | -1. | -4. | -4. | 011 |
| 42 | CIT | 0.93 | 0.62 | 1.86 | 1.5 | -2. | -3. | 011 |
| 43 | CLK1 | 0.62 | 0.31 | 2.17 | 2. | -3.5 | -7. | 011 |
| 44 | CLK2 | 0.93 | 1.55 | 7.13 | -1.7 | -7.7 | -4.6 | 011 |
| 45 | CLK3 | 0.62 | 0.62 | 5.58 | -1. | -9. | -9. | 011 |
| 46 | CLK4 | 0.31 | 0. | 1.55 | 100. | -5. | -500. | 011 |
| 47 | COL4A3BP | 2.17 | 1.86 | 4.96 | 1.2 | -2.3 | -2.7 | 011 |
| 48 | CPNE3 | 4.34 | 4.03 | 15.5 | 1.1 | -3.6 | -3.8 | 011 |
| 49 | CRKRS | 1.55 | 0.93 | 2.79 | 1.7 | -1.8 | -3. | 011 |
| 50 | CSK | 2.48 | 3.1 | 13.64 | -1.2 | -5.5 | -4.4 | 011 |
| 51 | CSNK1G2 | 0.62 | 0.93 | 1.86 | -1.5 | -3. | -2. | 011 |
| 52 | CSNK2A2 | 0.62 | 0.93 | 4.34 | -1.5 | -7. | -4.7 | 011 |
| 53 | DAPK3 | 0.93 | 1.24 | 2.79 | -1.3 | -3. | -2.2 | 011 |
| 54 | DDR1 | 1.24 | 1.24 | 9.3 | -1. | -7.5 | -7.5 | 011 |
| 55 | DYRK1A | 1.24 | 1.24 | 6.51 | -1. | -5.2 | -5.2 | 011 |
| 56 | DYRK1B | 0.62 | 0.31 | 1.55 | 2. | -2.5 | -5. | 011 |
| 57 | DYRK2 | 1.55 | 1.55 | 8.99 | -1. | -5.8 | -5.8 | 011 |
| 58 | EIF2AK3 | 0.93 | 0.31 | 1.86 | 3. | -2. | -6. | 011 |
| 59 | ERBB2 | 1.24 | 0.93 | 8.99 | 1.3 | -7.2 | -9.7 | 011 |
| 60 | ERBB3 | 1.55 | 1.55 | 19.84 | -1. | -12.8 | -12.8 | 011 |
| 61 | ERN1 | 0.62 | 0.31 | 5.27 | 2. | -8.5 | -17. | 011 |
| 62 | ERN2 | 0. | 0. | 9.92 | -1. | -3200. | -3200. | 011 |
| 63 | FRK | 0. | 0. | 2.48 | -1. | -800. | -800. | 011 |
| 64 | GAK | 0.62 | 0.93 | 6.82 | -1.5 | -11. | -7.3 | 011 |
| 65 | HIPK1 | 0.62 | 0.62 | 1.86 | -1. | -3. | -3. | 011 |
| 66 | HSPB8 | 0.62 | 0. | 4.03 | 200. | -6.5 | -1300. | 011 |
| 67 | HUS1 | 2.48 | 1.86 | 7.75 | 1.3 | -3.1 | -4.2 | 011 |
| 68 | IKBKB | 0.31 | 0.31 | 1.24 | -1. | -4. | -4. | 011 |
| 69 | IKBKE | 0. | 0. | 1.86 | -1. | -600. | -600. | 011 |
| 70 | ILK | 1.24 | 0.62 | 2.48 | 2. | -2. | -4. | 011 |
| 71 | INSR | 0.62 | 0. | 2.17 | 200. | -3.5 | -700. | 011 |
| 72 | IRAK1 | 5.58 | 4.96 | 12.71 | 1.1 | -2.3 | -2.6 | 011 |
| 73 | IRAK2 | 0. | 0.31 | 10.23 | -100. | -3300. | -33. | 011 |
| 74 | KIAA0971 | 1.24 | 0.93 | 2.48 | 1.3 | -2. | -2.7 | 011 |
| 75 | LATS1 | 1.86 | 1.55 | 8.06 | 1.2 | -4.3 | -5.2 | 011 |
| 76 | LIMK2 | 0.31 | 0.62 | 8.06 | -2. | -26. | -13. | 011 |
| 77 | LYN | 0.31 | 0.62 | 2.17 | -2. | -7. | -3.5 | 011 |
| 78 | MAP2K3 | 1.55 | 0.93 | 10.85 | 1.7 | -7. | -11.7 | 011 |
| 79 | MAP2K5 | 0.31 | 0. | 1.86 | 100. | -6. | -600. | 011 |
| 80 | MAP2K7 | 0.31 | 0.62 | 2.79 | -2. | -9. | -4.5 | 011 |
| 81 | MAP3K1 | 0.31 | 0.93 | 3.72 | -3. | -12. | -4. | 011 |
| 82 | MAP3K10 | 0. | 0. | 0.93 | -1. | -300. | -300. | 011 |

Continued on next page

**Table B.5 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 83 | MAP3K4 | 0.62 | 0.93 | 3.41 | -1.5 | -5.5 | -3.7 | 011 |
| 84 | MAP3K5 | 0. | 0. | 2.17 | -1. | -700. | -700. | 011 |
| 85 | MAP4K3 | 1.24 | 0.62 | 4.03 | 2. | -3.2 | -6.5 | 011 |
| 86 | MAPK12 | 0. | 0.31 | 1.24 | -100. | -400. | -4. | 011 |
| 87 | MAPK14 | 2.48 | 2.79 | 9.3 | -1.1 | -3.8 | -3.3 | 011 |
| 88 | MAPK3 | 2.17 | 2.17 | 8.68 | -1. | -4. | -4. | 011 |
| 89 | MAPK8 | 1.55 | 0.93 | 8.37 | 1.7 | -5.4 | -9. | 011 |
| 90 | MAPKAPK2 | 1.86 | 1.55 | 6.51 | 1.2 | -3.5 | -4.2 | 011 |
| 91 | MAPKAPK5 | 2.17 | 2.17 | 9.92 | -1. | -4.6 | -4.6 | 011 |
| 92 | MARK1 | 0. | 0. | 0.93 | -1. | -300. | -300. | 011 |
| 93 | MARK4 | 0.31 | 0.93 | 2.17 | -3. | -7. | -2.3 | 011 |
| 94 | MAST4 | 0. | 0. | 3.1 | -1. | -1000. | -1000. | 011 |
| 95 | MGC16169 | 0.62 | 0.62 | 1.55 | -1. | -2.5 | -2.5 | 011 |
| 96 | MKNK1 | 0.93 | 0.62 | 5.89 | 1.5 | -6.3 | -9.5 | 011 |
| 97 | MTOR | 0.31 | 0.31 | 1.24 | -1. | -4. | -4. | 011 |
| 98 | NEK11 | 0. | 0. | 0.93 | -1. | -300. | -300. | 011 |
| 99 | NEK2 | 0.62 | 0.62 | 2.48 | -1. | -4. | -4. | 011 |
| 100 | NEK3 | 0.62 | 0.31 | 1.55 | 2. | -2.5 | -5. | 011 |
| 101 | NEK4 | 0.31 | 0.62 | 1.86 | -2. | -6. | -3. | 011 |
| 102 | NEK7 | 2.79 | 3.1 | 7.75 | -1.1 | -2.8 | -2.5 | 011 |
| 103 | NEK9 | 1.24 | 0.93 | 3.41 | 1.3 | -2.8 | -3.7 | 011 |
| 104 | OXSR1 | 1.55 | 1.86 | 7.44 | -1.2 | -4.8 | -4. | 011 |
| 105 | PAK1 | 2.17 | 2.79 | 7.44 | -1.3 | -3.4 | -2.7 | 011 |
| 106 | PAK2 | 1.55 | 1.86 | 5.89 | -1.2 | -3.8 | -3.2 | 011 |
| 107 | PAN3 | 0.62 | 0.62 | 2.79 | -1. | -4.5 | -4.5 | 011 |
| 108 | PCTK2 | 0.93 | 0.93 | 9.3 | -1. | -10. | -10. | 011 |
| 109 | PDIK1L | 0.31 | 0.93 | 3.1 | -3. | -10. | -3.3 | 011 |
| 110 | PDPK1 | 1.55 | 1.24 | 4.03 | 1.2 | -2.6 | -3.2 | 011 |
| 111 | PKN1 | 2.48 | 1.86 | 13.33 | 1.3 | -5.4 | -7.2 | 011 |
| 112 | PKN2 | 1.55 | 1.55 | 7.44 | -1. | -4.8 | -4.8 | 011 |
| 113 | PLK3 | 0. | 0.31 | 3.1 | -100. | -1000. | -10. | 011 |
| 114 | PRKAA1 | 3.72 | 3.1 | 9.3 | 1.2 | -2.5 | -3. | 011 |
| 115 | PRKAA2 | 0.93 | 0.62 | 4.34 | 1.5 | -4.7 | -7. | 011 |
| 116 | PRKACA | 4.96 | 4.96 | 11.16 | -1. | -2.2 | -2.2 | 011 |
| 117 | PRKCE | 0.31 | 0.31 | 1.24 | -1. | -4. | -4. | 011 |
| 118 | PRKCI | 2.79 | 2.48 | 12.71 | 1.1 | -4.6 | -5.1 | 011 |
| 119 | PRKCZ | 0.31 | 0.93 | 3.1 | -3. | -10. | -3.3 | 011 |
| 120 | PRKD2 | 0.62 | 0.62 | 5.27 | -1. | -8.5 | -8.5 | 011 |
| 121 | PRKD3 | 1.24 | 0.62 | 4.03 | 2. | -3.2 | -6.5 | 011 |
| 122 | PRKDC | 10.23 | 10.85 | 14.88 | -1.1 | -1.5 | -1.4 | 011 |
| 123 | PRKX | 0. | 0. | 1.55 | -1. | -500. | -500. | 011 |
| 124 | PTK6 | 0. | 0.62 | 9.61 | -200. | -3100. | -15.5 | 011 |

Continued on next page

**Table B.5 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 125 | PXK | 0. | 0.62 | 2.79 | -200. | -900. | -4.5 | 011 |
| 126 | RIOK2 | 1.86 | 1.24 | 5.27 | 1.5 | -2.8 | -4.2 | 011 |
| 127 | RIOK3 | 0.62 | 1.24 | 7.75 | -2. | -12.5 | -6.2 | 011 |
| 128 | RIPK1 | 1.55 | 1.24 | 2.79 | 1.2 | -1.8 | -2.2 | 011 |
| 129 | RIPK4 | 1.86 | 2.17 | 8.99 | -1.2 | -4.8 | -4.1 | 011 |
| 130 | RNASEL | 0.31 | 0. | 1.86 | 100. | -6. | -600. | 011 |
| 131 | RPS6KB1 | 2.17 | 1.86 | 7.13 | 1.2 | -3.3 | -3.8 | 011 |
| 132 | RYK | 0.31 | 0.62 | 1.55 | -2. | -5. | -2.5 | 011 |
| 133 | SMG1 | 2.79 | 2.17 | 5.58 | 1.3 | -2. | -2.6 | 011 |
| 134 | SNF1LK2 | 0.93 | 1.24 | 3.72 | -1.3 | -4. | -3. | 011 |
| 135 | SNRK | 0.31 | 0.62 | 9.61 | -2. | -31. | -15.5 | 011 |
| 136 | STK16 | 0.62 | 0.31 | 2.48 | 2. | -4. | -8. | 011 |
| 137 | STK25 | 1.55 | 0.93 | 4.34 | 1.7 | -2.8 | -4.7 | 011 |
| 138 | STK32C | 0.31 | 0. | 1.86 | 100. | -6. | -600. | 011 |
| 139 | STK35 | 1.24 | 1.24 | 4.34 | -1. | -3.5 | -3.5 | 011 |
| 140 | STK38L | 0.93 | 1.24 | 4.34 | -1.3 | -4.7 | -3.5 | 011 |
| 141 | STYK1 | 0. | 0. | 0.93 | -1. | -300. | -300. | 011 |
| 142 | SYK | 0. | 0.62 | 7.44 | -200. | -2400. | -12. | 011 |
| 143 | TAF1 | 0.62 | 0.62 | 2.48 | -1. | -4. | -4. | 011 |
| 144 | TAF1L | 0.62 | 0.62 | 2.48 | -1. | -4. | -4. | 011 |
| 145 | TAOK3 | 0.93 | 0.31 | 8.06 | 3. | -8.7 | -26. | 011 |
| 146 | TESK1 | 0.31 | 0.31 | 1.24 | -1. | -4. | -4. | 011 |
| 147 | TLK1 | 1.24 | 0.93 | 6.2 | 1.3 | -5. | -6.7 | 011 |
| 148 | TLK2 | 1.55 | 1.86 | 4.34 | -1.2 | -2.8 | -2.3 | 011 |
| 149 | TNK1 | 0. | 0. | 1.24 | -1. | -400. | -400. | 011 |
| 150 | TRIB3 | 1.24 | 0.62 | 12.71 | 2. | -10.2 | -20.5 | 011 |
| 151 | TRIO | 3.72 | 3.72 | 5.58 | -1. | -1.5 | -1.5 | 011 |
| 152 | TSSK4 | 0. | 0. | 0.93 | -1. | -300. | -300. | 011 |
| 153 | UHMK1 | 4.34 | 4.65 | 20.46 | -1.1 | -4.7 | -4.4 | 011 |
| 154 | VRK1 | 2.79 | 2.79 | 9.3 | -1. | -3.3 | -3.3 | 011 |
| 155 | VRK3 | 1.24 | 0.93 | 3.72 | 1.3 | -3. | -4. | 011 |
| 156 | CHEK1 | 0.93 | 1.86 | 1.24 | -2. | -1.3 | 1.5 | 100 |
| 157 | ICK | 0. | 0.93 | 0.62 | -300. | -200. | 1.5 | 100 |
| 158 | TAOK2 | 1.55 | 0.62 | 1.24 | 2.5 | 1.2 | -2. | 100 |
| 159 | ALS2CR2 | 1.24 | 0. | 1.24 | 400. | -1. | -400. | 101 |
| 160 | BMPR2 | 2.17 | 0.62 | 1.55 | 3.5 | 1.4 | -2.5 | 101 |
| 161 | FYN | 0. | 0.93 | 0. | -300. | -1. | 300. | 101 |
| 162 | JAK1 | 4.34 | 0.93 | 4.96 | 4.7 | -1.1 | -5.3 | 101 |
| 163 | LRRK1 | 0.93 | 0. | 1.24 | 300. | -1.3 | -400. | 101 |
| 164 | MAP3K8 | 1.24 | 0. | 1.86 | 400. | -1.5 | -600. | 101 |
| 165 | MELK | 9.3 | 4.03 | 9.92 | 2.3 | -1.1 | -2.5 | 101 |
| 166 | MERTK | 0. | 0.93 | 0. | -300. | -1. | 300. | 101 |

Continued on next page

**Table B.5 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 167 | NRBP1 | 13.02 | 3.72 | 13.64 | 3.5 | -1. | -3.7 | 101 |
| 168 | PRKCA | 2.79 | 1.24 | 2.17 | 2.2 | 1.3 | -1.8 | 101 |
| 169 | RPS6KC1 | 1.55 | 0.62 | 2.17 | 2.5 | -1.4 | -3.5 | 101 |
| 170 | ULK2 | 1.24 | 0. | 1.24 | 400. | -1. | -400. | 101 |
| 171 | AKT3 | 1.55 | 0. | 0. | 500. | 500. | -1. | 110 |
| 172 | AXL | 8.99 | 2.48 | 1.86 | 3.6 | 4.8 | 1.3 | 110 |
| 173 | CAMK1D | 1.55 | 0.62 | 0.31 | 2.5 | 5. | 2. | 110 |
| 174 | FGFR1 | 7.13 | 0. | 0. | 2300. | 2300. | -1. | 110 |
| 175 | FGFR4 | 1.55 | 0. | 0. | 500. | 500. | -1. | 110 |
| 176 | HAK | 1.24 | 0. | 0. | 400. | 400. | -1. | 110 |
| 177 | LOC91461 | 4.96 | 0. | 0. | 1600. | 1600. | -1. | 110 |
| 178 | MAP3K14 | 1.55 | 0. | 0.62 | 500. | 2.5 | -200. | 110 |
| 179 | MAP3K9 | 0. | 0.93 | 1.55 | -300. | -500. | -1.7 | 110 |
| 180 | NTRK3 | 4.03 | 0. | 0. | 1300. | 1300. | -1. | 110 |
| 181 | NUAK2 | 4.03 | 0.31 | 0.93 | 13. | 4.3 | -3. | 110 |
| 182 | TGFBR1 | 2.79 | 0.31 | 0.93 | 9. | 3. | -3. | 110 |
| 183 | ABL1 | 2.17 | 0.93 | 5.27 | 2.3 | -2.4 | -5.7 | 111 |
| 184 | ADCK2 | 2.48 | 0.93 | 7.13 | 2.7 | -2.9 | -7.7 | 111 |
| 185 | ADRBK1 | 2.17 | 9.92 | 15.19 | -4.6 | -7. | -1.5 | 111 |
| 186 | AKT2 | 10.85 | 4.65 | 17.05 | 2.3 | -1.6 | -3.7 | 111 |
| 187 | ARAF | 3.1 | 1.86 | 5.89 | 1.7 | -1.9 | -3.2 | 111 |
| 188 | AURKB | 12.09 | 10.54 | 15.19 | 1.1 | -1.3 | -1.4 | 111 |
| 189 | BCR | 0.62 | 2.48 | 11.16 | -4. | -18. | -4.5 | 111 |
| 190 | BMPR1A | 3.1 | 1.24 | 6.82 | 2.5 | -2.2 | -5.5 | 111 |
| 191 | BMPR1B | 1.24 | 0.31 | 2.17 | 4. | -1.8 | -7. | 111 |
| 192 | BRD2 | 2.17 | 3.1 | 8.06 | -1.4 | -3.7 | -2.6 | 111 |
| 193 | BUB1B | 3.1 | 5.58 | 6.82 | -1.8 | -2.2 | -1.2 | 111 |
| 194 | CAMK2D | 3.1 | 4.96 | 5.89 | -1.6 | -1.9 | -1.2 | 111 |
| 195 | CAMKK1 | 1.24 | 0. | 2.48 | 400. | -2. | -800. | 111 |
| 196 | CDC2 | 26.35 | 17.36 | 63.24 | 1.5 | -2.4 | -3.6 | 111 |
| 197 | CDC42BPB | 1.86 | 0.31 | 5.58 | 6. | -3. | -18. | 111 |
| 198 | CDK4 | 12.71 | 14.57 | 23.25 | -1.1 | -1.8 | -1.6 | 111 |
| 199 | CDK6 | 3.1 | 1.86 | 11.16 | 1.7 | -3.6 | -6. | 111 |
| 200 | CHUK | 3.72 | 1.86 | 11.47 | 2. | -3.1 | -6.2 | 111 |
| 201 | CSNK1A1 | 1.86 | 3.1 | 15.5 | -1.7 | -8.3 | -5. | 111 |
| 202 | CSNK1D | 15.19 | 17.67 | 31.62 | -1.2 | -2.1 | -1.8 | 111 |
| 203 | CSNK1E | 3.41 | 5.58 | 60.76 | -1.6 | -17.8 | -10.9 | 111 |
| 204 | CSNK1G1 | 1.86 | 0.62 | 4.96 | 3. | -2.7 | -8. | 111 |
| 205 | CSNK1G3 | 2.48 | 1.55 | 7.44 | 1.6 | -3. | -4.8 | 111 |
| 206 | CSNK2A1 | 4.65 | 3.1 | 14.26 | 1.5 | -3.1 | -4.6 | 111 |
| 207 | DAPK1 | 5.58 | 0. | 0.93 | 1800. | 6. | -300. | 111 |
| 208 | EGFR | 2.79 | 0.93 | 6.2 | 3. | -2.2 | -6.7 | 111 |

Continued on next page

**Table B.5 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 209 | EIF2AK1 | 9.92 | 5.27 | 22.63 | 1.9 | -2.3 | -4.3 | 111 |
| 210 | EIF2AK2 | 4.65 | 2.79 | 9.92 | 1.7 | -2.1 | -3.6 | 111 |
| 211 | EIF2AK4 | 2.48 | 4.34 | 9.3 | -1.8 | -3.8 | -2.1 | 111 |
| 212 | EPHA2 | 1.55 | 3.72 | 13.64 | -2.4 | -8.8 | -3.7 | 111 |
| 213 | EPHB4 | 0. | 0.93 | 4.03 | -300. | -1300. | -4.3 | 111 |
| 214 | FGFRL1 | 0. | 1.24 | 2.48 | -400. | -800. | -2. | 111 |
| 215 | FLJ13149 | 3.1 | 0.93 | 7.13 | 3.3 | -2.3 | -7.7 | 111 |
| 216 | FLJ21901 | 2.79 | 0.93 | 5.58 | 3. | -2. | -6. | 111 |
| 217 | FLJ23356 | 1.86 | 3.72 | 9.92 | -2. | -5.3 | -2.7 | 111 |
| 218 | GRK6 | 4.03 | 1.86 | 8.37 | 2.2 | -2.1 | -4.5 | 111 |
| 219 | GSG2 | 1.55 | 0.62 | 3.1 | 2.5 | -2. | -5. | 111 |
| 220 | GSK3A | 6.82 | 2.17 | 15.19 | 3.1 | -2.2 | -7. | 111 |
| 221 | GSK3B | 5.58 | 4.34 | 13.33 | 1.3 | -2.4 | -3.1 | 111 |
| 222 | HIPK3 | 1.86 | 0.93 | 6.2 | 2. | -3.3 | -6.7 | 111 |
| 223 | IGF1R | 4.65 | 0.62 | 6.2 | 7.5 | -1.3 | -10. | 111 |
| 224 | KIAA1804 | 1.55 | 0. | 4.03 | 500. | -2.6 | -1300. | 111 |
| 225 | LIMK1 | 1.86 | 2.79 | 10.23 | -1.5 | -5.5 | -3.7 | 111 |
| 226 | LMTK2 | 2.17 | 0.93 | 11.47 | 2.3 | -5.3 | -12.3 | 111 |
| 227 | MAP2K1 | 7.44 | 9.92 | 30.38 | -1.3 | -4.1 | -3.1 | 111 |
| 228 | MAP2K2 | 9.92 | 4.96 | 23.25 | 2. | -2.3 | -4.7 | 111 |
| 229 | MAP2K4 | 2.79 | 1.86 | 6.51 | 1.5 | -2.3 | -3.5 | 111 |
| 230 | MAP3K2 | 2.48 | 1.55 | 4.65 | 1.6 | -1.9 | -3. | 111 |
| 231 | MAP3K7 | 2.17 | 0.93 | 5.89 | 2.3 | -2.7 | -6.3 | 111 |
| 232 | MAP4K4 | 4.65 | 8.37 | 14.26 | -1.8 | -3.1 | -1.7 | 111 |
| 233 | MAP4K5 | 2.48 | 5.89 | 10.54 | -2.4 | -4.2 | -1.8 | 111 |
| 234 | MAPK1 | 5.58 | 4.03 | 14.57 | 1.4 | -2.6 | -3.6 | 111 |
| 235 | MAPK13 | 0. | 9.3 | 23.25 | -3000. | -7500. | -2.5 | 111 |
| 236 | MAPK6 | 5.89 | 4.96 | 22.94 | 1.2 | -3.9 | -4.6 | 111 |
| 237 | MAPK9 | 2.48 | 4.96 | 8.06 | -2. | -3.2 | -1.6 | 111 |
| 238 | MAPKAPK3 | 0.31 | 1.24 | 3.1 | -4. | -10. | -2.5 | 111 |
| 239 | MARK2 | 3.1 | 5.27 | 13.95 | -1.7 | -4.5 | -2.6 | 111 |
| 240 | MASTL | 2.79 | 1.24 | 6.82 | 2.2 | -2.4 | -5.5 | 111 |
| 241 | MET | 15.5 | 10.54 | 16.43 | 1.5 | -1.1 | -1.6 | 111 |
| 242 | MINK1 | 2.17 | 0.62 | 8.68 | 3.5 | -4. | -14. | 111 |
| 243 | MKNK2 | 2.17 | 3.1 | 14.57 | -1.4 | -6.7 | -4.7 | 111 |
| 244 | MLKL | 2.48 | 0.93 | 10.85 | 2.7 | -4.4 | -11.7 | 111 |
| 245 | MST1R | 0.31 | 2.48 | 22.01 | -8. | -71. | -8.9 | 111 |
| 246 | MST4 | 1.86 | 2.79 | 11.78 | -1.5 | -6.3 | -4.2 | 111 |
| 247 | MYLK | 2.48 | 1.55 | 0. | 1.6 | 800. | 500. | 111 |
| 248 | PAK4 | 3.72 | 1.86 | 9.92 | 2. | -2.7 | -5.3 | 111 |
| 249 | PAK6 | 0. | 0.93 | 3.1 | -300. | -1000. | -3.3 | 111 |
| 250 | PBK | 8.06 | 5.58 | 10.23 | 1.4 | -1.3 | -1.8 | 111 |

Continued on next page

**Table B.5 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 251 | PCTK1 | 6.2 | 8.06 | 15.19 | -1.3 | -2.4 | -1.9 | 111 |
| 252 | PDK1 | 0.62 | 1.86 | 10.54 | -3. | -17. | -5.7 | 111 |
| 253 | PDK2 | 1.86 | 0.93 | 4.03 | 2. | -2.2 | -4.3 | 111 |
| 254 | PDK4 | 7.44 | 0. | 2.48 | 2400. | 3. | -800. | 111 |
| 255 | PIM1 | 1.86 | 2.79 | 9.3 | -1.5 | -5. | -3.3 | 111 |
| 256 | PIM3 | 0.31 | 1.55 | 8.37 | -5. | -27. | -5.4 | 111 |
| 257 | PLK1 | 4.96 | 3.72 | 8.68 | 1.3 | -1.8 | -2.3 | 111 |
| 258 | PLK2 | 11.47 | 4.34 | 51.15 | 2.6 | -4.5 | -11.8 | 111 |
| 259 | PRKACB | 1.55 | 2.48 | 5.27 | -1.6 | -3.4 | -2.1 | 111 |
| 260 | PRKCD | 1.24 | 2.17 | 17.36 | -1.8 | -14. | -8. | 111 |
| 261 | PRKCH | 0. | 1.55 | 7.13 | -500. | -2300. | -4.6 | 111 |
| 262 | PRPF4B | 4.03 | 3.1 | 8.37 | 1.3 | -2.1 | -2.7 | 111 |
| 263 | PTK2 | 7.13 | 8.68 | 14.26 | -1.2 | -2. | -1.6 | 111 |
| 264 | RAF1 | 4.03 | 2.79 | 12.4 | 1.4 | -3.1 | -4.4 | 111 |
| 265 | RIPK2 | 7.44 | 3.41 | 16.12 | 2.2 | -2.2 | -4.7 | 111 |
| 266 | ROCK1 | 2.79 | 1.24 | 6.51 | 2.2 | -2.3 | -5.2 | 111 |
| 267 | ROCK2 | 10.54 | 2.79 | 5.89 | 3.8 | 1.8 | -2.1 | 111 |
| 268 | RPS6KA1 | 1.55 | 2.48 | 8.99 | -1.6 | -5.8 | -3.6 | 111 |
| 269 | RPS6KA4 | 1.86 | 3.41 | 4.96 | -1.8 | -2.7 | -1.5 | 111 |
| 270 | RPS6KB2 | 0.93 | 2.17 | 3.1 | -2.3 | -3.3 | -1.4 | 111 |
| 271 | SCYL1 | 1.24 | 2.17 | 4.96 | -1.8 | -4. | -2.3 | 111 |
| 272 | SCYL2 | 3.1 | 2.17 | 10.54 | 1.4 | -3.4 | -4.9 | 111 |
| 273 | SGK | 9.92 | 0.31 | 2.79 | 32. | 3.6 | -9. | 111 |
| 274 | SLK | 2.48 | 1.24 | 19.53 | 2. | -7.9 | -15.8 | 111 |
| 275 | SNF1LK | 17.98 | 1.24 | 53.32 | 14.5 | -3. | -43. | 111 |
| 276 | SRC | 4.03 | 0.31 | 9.92 | 13. | -2.5 | -32. | 111 |
| 277 | SRPK1 | 4.03 | 5.89 | 19.84 | -1.5 | -4.9 | -3.4 | 111 |
| 278 | SRPK2 | 2.48 | 1.24 | 4.65 | 2. | -1.9 | -3.8 | 111 |
| 279 | STK17A | 2.79 | 6.51 | 25.42 | -2.3 | -9.1 | -3.9 | 111 |
| 280 | STK17B | 0.62 | 1.86 | 13.64 | -3. | -22. | -7.3 | 111 |
| 281 | STK24 | 4.65 | 5.89 | 23.87 | -1.3 | -5.1 | -4.1 | 111 |
| 282 | STK38 | 2.79 | 1.86 | 5.89 | 1.5 | -2.1 | -3.2 | 111 |
| 283 | STK39 | 2.79 | 4.03 | 10.54 | -1.4 | -3.8 | -2.6 | 111 |
| 284 | STK40 | 0.31 | 2.17 | 3.41 | -7. | -11. | -1.6 | 111 |
| 285 | TAOK1 | 4.65 | 2.17 | 11.78 | 2.1 | -2.5 | -5.4 | 111 |
| 286 | TBK1 | 2.48 | 1.24 | 8.06 | 2. | -3.2 | -6.5 | 111 |
| 287 | TBRG4 | 2.17 | 0.93 | 8.68 | 2.3 | -4. | -9.3 | 111 |
| 288 | TGFBR2 | 5.58 | 1.55 | 10.23 | 3.6 | -1.8 | -6.6 | 111 |
| 289 | TP53RK | 1.86 | 2.79 | 13.33 | -1.5 | -7.2 | -4.8 | 111 |
| 290 | TRIB1 | 1.24 | 2.48 | 5.89 | -2. | -4.8 | -2.4 | 111 |
| 291 | TRIB2 | 0. | 3.1 | 18.29 | -1000. | -5900. | -5.9 | 111 |
| 292 | TRPM7 | 1.55 | 0.62 | 3.41 | 2.5 | -2.2 | -5.5 | 111 |

Continued on next page

**Table B.5 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 293 | TTK | 2.48 | 3.72 | 7.13 | -1.5 | -2.9 | -1.9 | 111 |
| 294 | ULK1 | 1.86 | 0. | 4.03 | 600. | -2.2 | -1300. | 111 |
| 295 | ULK3 | 2.48 | 1.24 | 8.99 | 2. | -3.6 | -7.2 | 111 |
| 296 | VRK2 | 2.17 | 0.93 | 6.82 | 2.3 | -3.1 | -7.3 | 111 |
| 297 | WNK1 | 6.51 | 2.17 | 11.16 | 3. | -1.7 | -5.1 | 111 |
| 298 | YES1 | 4.65 | 6.82 | 16.74 | -1.5 | -3.6 | -2.5 | 111 |

Table B.6: List of significantly changed gene in the group: (1 - $H3255$, 2 - $H827$, 3 - $H1975$), along with pairwise fold changes and significant change markers

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 1 | AAK1 | 3.41 | 2.79 | 4.03 | 1.2 | -1.2 | -1.4 | 001 |
| 2 | CASK | 2.79 | 2.17 | 3.1 | 1.3 | -1.1 | -1.4 | 001 |
| 3 | CLK2 | 2.79 | 2.48 | 3.41 | 1.1 | -1.2 | -1.4 | 001 |
| 4 | ICK | 0.93 | 0.31 | 1.24 | 3. | -1.3 | -4. | 001 |
| 5 | MARK3 | 1.24 | 0.62 | 1.55 | 2. | -1.2 | -2.5 | 001 |
| 6 | NEK11 | 0.62 | 0. | 0.93 | 200. | -1.5 | -300. | 001 |
| 7 | PKN3 | 0.31 | 0.93 | 0. | -3. | 100. | 300. | 001 |
| 8 | PXK | 0.62 | 0. | 1.24 | 200. | -2. | -400. | 001 |
| 9 | SNRK | 0.62 | 0.31 | 1.24 | 2. | -2. | -4. | 001 |
| 10 | STK11 | 1.86 | 1.24 | 2.17 | 1.5 | -1.2 | -1.8 | 001 |
| 11 | TLK1 | 1.86 | 1.24 | 2.17 | 1.5 | -1.2 | -1.8 | 001 |
| 12 | ULK1 | 1.55 | 1.24 | 2.17 | 1.2 | -1.4 | -1.8 | 001 |
| 13 | ABL2 | 0.93 | 1.55 | 1.86 | -1.7 | -2. | -1.2 | 010 |
| 14 | CAMKK2 | 1.55 | 2.17 | 2.48 | -1.4 | -1.6 | -1.1 | 010 |
| 15 | CDK6 | 3.72 | 3.1 | 2.48 | 1.2 | 1.5 | 1.2 | 010 |
| 16 | CDK9 | 1.55 | 2.17 | 2.79 | -1.4 | -1.8 | -1.3 | 010 |
| 17 | HIPK1 | 1.55 | 0.93 | 0.31 | 1.7 | 5. | 3. | 010 |
| 18 | MAP3K13 | 0.93 | 0.62 | 0. | 1.5 | 300. | 200. | 010 |
| 19 | MARK1 | 0.93 | 0.31 | 0. | 3. | 300. | 100. | 010 |
| 20 | MGC16169 | 2.17 | 1.55 | 0.93 | 1.4 | 2.3 | 1.7 | 010 |
| 21 | MINK1 | 4.03 | 4.65 | 4.96 | -1.2 | -1.2 | -1.1 | 010 |
| 22 | NRK | 1.24 | 0.62 | 0. | 2. | 400. | 200. | 010 |
| 23 | PAN3 | 1.24 | 0.93 | 0.31 | 1.3 | 4. | 3. | 010 |
| 24 | PSKH1 | 0.93 | 1.55 | 2.17 | -1.7 | -2.3 | -1.4 | 010 |
| 25 | RIPK4 | 4.03 | 3.41 | 2.79 | 1.2 | 1.4 | 1.2 | 010 |
| 26 | RPS6KA5 | 1.24 | 0.62 | 0.31 | 2. | 4. | 2. | 010 |
| 27 | STK25 | 3.41 | 2.79 | 2.17 | 1.2 | 1.6 | 1.3 | 010 |
| 28 | AKT3 | 0. | 0. | 0.93 | -1. | -300. | -300. | 011 |
| 29 | ALS2CR2 | 1.24 | 0.93 | 2.17 | 1.3 | -1.8 | -2.3 | 011 |
| 30 | AXL | 0.93 | 0.93 | 8.99 | -1. | -9.7 | -9.7 | 011 |
| 31 | BMPR1A | 1.55 | 1.55 | 3.72 | -1. | -2.4 | -2.4 | 011 |
| 32 | BUB1B | 5.27 | 5.89 | 4.03 | -1.1 | 1.3 | 1.5 | 011 |
| 33 | CDK7 | 1.24 | 1.86 | 3.1 | -1.5 | -2.5 | -1.7 | 011 |
| 34 | CDK8 | 1.55 | 1.24 | 2.48 | 1.2 | -1.6 | -2. | 011 |
| 35 | CHEK1 | 0.62 | 1.24 | 2.17 | -2. | -3.5 | -1.8 | 011 |
| 36 | CLK1 | 2.48 | 2.17 | 0.62 | 1.1 | 4. | 3.5 | 011 |
| 37 | CSNK1G1 | 1.24 | 0.93 | 2.48 | 1.3 | -2. | -2.7 | 011 |
| 38 | CSNK2A1 | 3.1 | 3.41 | 4.96 | -1.1 | -1.6 | -1.5 | 011 |
| 39 | DYRK3 | 0.31 | 0. | 1.24 | 100. | -4. | -400. | 011 |
| 40 | GRK5 | 0.31 | 0.31 | 1.24 | -1. | -4. | -4. | 011 |

Continued on next page

**Table B.6 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 41 | HSPB8 | 0.31 | 0.62 | 3.72 | -2. | -12. | -6. | 011 |
| 42 | IRAK4 | 0. | 0.31 | 1.24 | -100. | -400. | -4. | 011 |
| 43 | MAP3K2 | 2.48 | 2.79 | 3.72 | -1.1 | -1.5 | -1.3 | 011 |
| 44 | MAP3K9 | 2.17 | 1.86 | 0.62 | 1.2 | 3.5 | 3. | 011 |
| 45 | MAP4K3 | 3.41 | 2.79 | 4.96 | 1.2 | -1.5 | -1.8 | 011 |
| 46 | MAP4K5 | 7.44 | 8.06 | 4.96 | -1.1 | 1.5 | 1.6 | 011 |
| 47 | MAST2 | 3.41 | 3.1 | 1.86 | 1.1 | 1.8 | 1.7 | 011 |
| 48 | MASTL | 2.17 | 2.79 | 1.24 | -1.3 | 1.8 | 2.2 | 011 |
| 49 | MKNK1 | 3.1 | 2.79 | 0.93 | 1.1 | 3.3 | 3. | 011 |
| 50 | NUAK2 | 2.17 | 2.48 | 1.24 | -1.1 | 1.8 | 2. | 011 |
| 51 | OXSR1 | 1.24 | 1.55 | 2.48 | -1.2 | -2. | -1.6 | 011 |
| 52 | PFTK1 | 0.31 | 0.31 | 1.24 | -1. | -4. | -4. | 011 |
| 53 | PKN1 | 4.65 | 4.34 | 2.79 | 1.1 | 1.7 | 1.6 | 011 |
| 54 | PRKAA2 | 1.55 | 1.86 | 0. | -1.2 | 500. | 600. | 011 |
| 55 | RPS6KA4 | 3.1 | 3.41 | 4.34 | -1.1 | -1.4 | -1.3 | 011 |
| 56 | RPS6KC1 | 1.86 | 1.24 | 3.41 | 1.5 | -1.8 | -2.8 | 011 |
| 57 | SLK | 1.86 | 2.48 | 4.65 | -1.3 | -2.5 | -1.9 | 011 |
| 58 | SNF1LK2 | 0.93 | 1.55 | 2.48 | -1.7 | -2.7 | -1.6 | 011 |
| 59 | STK32C | 0. | 0. | 1.24 | -1. | -400. | -400. | 011 |
| 60 | STK35 | 1.55 | 1.55 | 3.72 | -1. | -2.4 | -2.4 | 011 |
| 61 | SYK | 1.24 | 0.93 | 0. | 1.3 | 400. | 300. | 011 |
| 62 | TTK | 4.96 | 5.27 | 4.03 | -1.1 | 1.2 | 1.3 | 011 |
| 63 | CAMK1 | 1.24 | 0.31 | 0.93 | 4. | 1.3 | -3. | 100 |
| 64 | CDKL2 | 0.93 | 0. | 0.31 | 300. | 3. | -100. | 100 |
| 65 | CSNK1A1 | 5.58 | 4.65 | 4.96 | 1.2 | 1.1 | -1.1 | 100 |
| 66 | FASTK | 0.93 | 0. | 0.62 | 300. | 1.5 | -200. | 100 |
| 67 | GSG2 | 1.55 | 2.79 | 2.17 | -1.8 | -1.4 | 1.3 | 100 |
| 68 | MAP4K2 | 0.93 | 0. | 0.31 | 300. | 3. | -100. | 100 |
| 69 | MAPKAPK5 | 4.03 | 3.1 | 3.72 | 1.3 | 1.1 | -1.2 | 100 |
| 70 | MGC5297 | 1.24 | 0.31 | 0.62 | 4. | 2. | -2. | 100 |
| 71 | MYO3B | 0.93 | 0. | 0.62 | 300. | 1.5 | -200. | 100 |
| 72 | PRKACB | 0.62 | 1.86 | 1.24 | -3. | -2. | 1.5 | 100 |
| 73 | PTK6 | 0.93 | 0. | 0.31 | 300. | 3. | -100. | 100 |
| 74 | ROCK1 | 4.03 | 2.79 | 3.41 | 1.4 | 1.2 | -1.2 | 100 |
| 75 | TESK1 | 1.24 | 0.31 | 0.93 | 4. | 1.3 | -3. | 100 |
| 76 | TYK2 | 1.24 | 0.31 | 0.62 | 4. | 2. | -2. | 100 |
| 77 | TYRO3 | 0.93 | 0. | 0.31 | 300. | 3. | -100. | 100 |
| 78 | BMP2K | 2.48 | 1.55 | 2.79 | 1.6 | -1.1 | -1.8 | 101 |
| 79 | BRDT | 0. | 1.55 | 0. | -500. | -1. | 500. | 101 |
| 80 | BUB1 | 3.72 | 5.58 | 3.1 | -1.5 | 1.2 | 1.8 | 101 |
| 81 | CAMK2D | 8.68 | 3.72 | 8.06 | 2.3 | 1.1 | -2.2 | 101 |
| 82 | CDC2L5 | 4.34 | 2.48 | 4.03 | 1.8 | 1.1 | -1.6 | 101 |

Continued on next page

**Table B.6 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 83 | CDC42BPB | 4.34 | 2.48 | 3.72 | 1.8 | 1.2 | -1.5 | 101 |
| 84 | CDK10 | 5.58 | 6.82 | 5.89 | -1.2 | -1.1 | 1.2 | 101 |
| 85 | CDK2 | 2.79 | 1.24 | 2.48 | 2.2 | 1.1 | -2. | 101 |
| 86 | CDK5 | 3.41 | 0.62 | 2.79 | 5.5 | 1.2 | -4.5 | 101 |
| 87 | COL4A3BP | 3.72 | 4.96 | 3.72 | -1.3 | -1. | 1.3 | 101 |
| 88 | CSK | 2.79 | 1.24 | 2.79 | 2.2 | -1. | -2.2 | 101 |
| 89 | CSNK1G2 | 2.48 | 0.93 | 2.48 | 2.7 | -1. | -2.7 | 101 |
| 90 | CSNK1G3 | 3.41 | 2.48 | 3.72 | 1.4 | -1.1 | -1.5 | 101 |
| 91 | FER | 1.55 | 0.62 | 1.55 | 2.5 | -1. | -2.5 | 101 |
| 92 | GSK3B | 12.09 | 8.06 | 11.78 | 1.5 | 1. | -1.5 | 101 |
| 93 | LIMK1 | 21.39 | 8.06 | 20.77 | 2.7 | 1. | -2.6 | 101 |
| 94 | LIMK2 | 1.86 | 0.62 | 2.17 | 3. | -1.2 | -3.5 | 101 |
| 95 | MAP2K7 | 1.24 | 2.17 | 1.24 | -1.8 | -1. | 1.8 | 101 |
| 96 | MAP3K3 | 2.17 | 1.24 | 2.79 | 1.8 | -1.3 | -2.2 | 101 |
| 97 | MAPK7 | 2.79 | 1.24 | 2.17 | 2.2 | 1.3 | -1.8 | 101 |
| 98 | MAPKAPK2 | 4.03 | 2.79 | 4.34 | 1.4 | -1.1 | -1.6 | 101 |
| 99 | NEK2 | 0.62 | 2.48 | 1.24 | -4. | -2. | 2. | 101 |
| 100 | PAK6 | 4.96 | 1.24 | 4.96 | 4. | -1. | -4. | 101 |
| 101 | PCTK2 | 3.1 | 1.55 | 2.48 | 2. | 1.2 | -1.6 | 101 |
| 102 | PDPK1 | 3.1 | 0.93 | 3.1 | 3.3 | -1. | -3.3 | 101 |
| 103 | PHKG2 | 2.17 | 0.93 | 1.86 | 2.3 | 1.2 | -2. | 101 |
| 104 | PRKDC | 6.2 | 5.27 | 6.82 | 1.2 | -1.1 | -1.3 | 101 |
| 105 | RPS6KA3 | 0.93 | 1.86 | 0.93 | -2. | -1. | 2. | 101 |
| 106 | RPS6KB1 | 2.48 | 1.24 | 2.48 | 2. | -1. | -2. | 101 |
| 107 | SCYL1 | 4.03 | 2.79 | 4.03 | 1.4 | -1. | -1.4 | 101 |
| 108 | SRC | 1.55 | 2.79 | 0.93 | -1.8 | 1.7 | 3. | 101 |
| 109 | TBK1 | 3.1 | 32.24 | 3.41 | -10.4 | -1.1 | 9.5 | 101 |
| 110 | TBRG4 | 3.41 | 1.55 | 3.41 | 2.2 | -1. | -2.2 | 101 |
| 111 | TLK2 | 2.17 | 1.24 | 2.17 | 1.8 | -1. | -1.8 | 101 |
| 112 | AURKB | 10.23 | 15.5 | 14.88 | -1.5 | -1.5 | 1. | 110 |
| 113 | CDC2L6 | 2.48 | 1.55 | 1.24 | 1.6 | 2. | 1.2 | 110 |
| 114 | CDKL5 | 1.24 | 0.31 | 0. | 4. | 400. | 100. | 110 |
| 115 | CHEK2 | 2.79 | 0.93 | 1.24 | 3. | 2.2 | -1.3 | 110 |
| 116 | CRKRS | 3.1 | 1.24 | 1.55 | 2.5 | 2. | -1.2 | 110 |
| 117 | CSNK2A2 | 0.31 | 1.55 | 1.55 | -5. | -5. | -1. | 110 |
| 118 | DAPK2 | 1.24 | 0. | 0.31 | 400. | 4. | -100. | 110 |
| 119 | DYRK1A | 3.72 | 1.24 | 1.24 | 3. | 3. | -1. | 110 |
| 120 | EIF2AK3 | 2.17 | 0.93 | 1.24 | 2.3 | 1.8 | -1.3 | 110 |
| 121 | EPHA1 | 2.48 | 0.31 | 0.62 | 8. | 4. | -2. | 110 |
| 122 | EPHA2 | 9.92 | 2.48 | 2.17 | 4. | 4.6 | 1.1 | 110 |
| 123 | EPHA4 | 7.75 | 0. | 0.31 | 2500. | 25. | -100. | 110 |
| 124 | EPHB3 | 2.17 | 0. | 0. | 700. | 700. | -1. | 110 |

Continued on next page

**Table B.6 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 125 | ERN1 | 3.41 | 0.93 | 0.93 | 3.7 | 3.7 | -1. | 110 |
| 126 | FGFR2 | 1.55 | 0. | 0. | 500. | 500. | -1. | 110 |
| 127 | FGFRL1 | 1.24 | 0.31 | 0.31 | 4. | 4. | -1. | 110 |
| 128 | FLJ21901 | 1.24 | 2.17 | 2.17 | -1.8 | -1.8 | -1. | 110 |
| 129 | FLJ23356 | 9.92 | 2.79 | 2.17 | 3.6 | 4.6 | 1.3 | 110 |
| 130 | FRK | 1.86 | 0.93 | 0.31 | 2. | 6. | 3. | 110 |
| 131 | FYN | 6.82 | 1.55 | 1.86 | 4.4 | 3.7 | -1.2 | 110 |
| 132 | GAK | 3.1 | 1.86 | 2.17 | 1.7 | 1.4 | -1.2 | 110 |
| 133 | GSK3A | 7.44 | 1.86 | 1.86 | 4. | 4. | -1. | 110 |
| 134 | IGF1R | 2.48 | 0.62 | 0.62 | 4. | 4. | -1. | 110 |
| 135 | IKBKB | 3.72 | 0. | 0.31 | 1200. | 12. | -100. | 110 |
| 136 | IRAK2 | 0. | 0.93 | 1.24 | -300. | -400. | -1.3 | 110 |
| 137 | KDR | 0.93 | 0. | 0. | 300. | 300. | -1. | 110 |
| 138 | LATS1 | 3.72 | 2.17 | 2.17 | 1.7 | 1.7 | -1. | 110 |
| 139 | LMTK2 | 4.34 | 2.79 | 3.1 | 1.6 | 1.4 | -1.1 | 110 |
| 140 | MAP2K3 | 4.34 | 3.1 | 3.1 | 1.4 | 1.4 | -1. | 110 |
| 141 | MAP3K5 | 3.1 | 0.31 | 0.93 | 10. | 3.3 | -3. | 110 |
| 142 | MAP3K6 | 1.24 | 0. | 0. | 400. | 400. | -1. | 110 |
| 143 | MAP4K4 | 2.48 | 11.78 | 12.4 | -4.8 | -5. | -1.1 | 110 |
| 144 | MAPK12 | 4.03 | 0.31 | 0.93 | 13. | 4.3 | -3. | 110 |
| 145 | MAPK9 | 4.34 | 2.17 | 2.79 | 2. | 1.6 | -1.3 | 110 |
| 146 | MARK2 | 8.37 | 5.27 | 5.89 | 1.6 | 1.4 | -1.1 | 110 |
| 147 | MARK4 | 1.86 | 0.93 | 0.62 | 2. | 3. | 1.5 | 110 |
| 148 | MLKL | 0.62 | 2.17 | 2.17 | -3.5 | -3.5 | -1. | 110 |
| 149 | MTOR | 1.55 | 0.31 | 0.31 | 5. | 5. | -1. | 110 |
| 150 | NEK4 | 1.24 | 0.31 | 0.31 | 4. | 4. | -1. | 110 |
| 151 | NLK | 1.86 | 0.62 | 0.93 | 3. | 2. | -1.5 | 110 |
| 152 | NRBP2 | 1.55 | 0.31 | 0.31 | 5. | 5. | -1. | 110 |
| 153 | PASK | 0.93 | 0. | 0. | 300. | 300. | -1. | 110 |
| 154 | PBK | 0.93 | 4.03 | 4.03 | -4.3 | -4.3 | -1. | 110 |
| 155 | PDIK1L | 2.17 | 1.24 | 0.62 | 1.8 | 3.5 | 2. | 110 |
| 156 | PIM1 | 9.92 | 3.72 | 3.72 | 2.7 | 2.7 | -1. | 110 |
| 157 | PIM3 | 5.27 | 0.93 | 1.24 | 5.7 | 4.2 | -1.3 | 110 |
| 158 | PKMYT1 | 2.17 | 0.62 | 1.24 | 3.5 | 1.8 | -2. | 110 |
| 159 | PLK2 | 26.97 | 7.75 | 8.06 | 3.5 | 3.3 | -1. | 110 |
| 160 | PRKAA1 | 12.4 | 5.58 | 5.89 | 2.2 | 2.1 | -1.1 | 110 |
| 161 | PRKCH | 4.65 | 1.55 | 1.86 | 3. | 2.5 | -1.2 | 110 |
| 162 | PRKCI | 34.41 | 5.58 | 6.2 | 6.2 | 5.5 | -1.1 | 110 |
| 163 | PRKD2 | 3.1 | 0.62 | 0.62 | 5. | 5. | -1. | 110 |
| 164 | PTK7 | 5.58 | 0.93 | 0.62 | 6. | 9. | 1.5 | 110 |
| 165 | RIOK1 | 0.93 | 0. | 0. | 300. | 300. | -1. | 110 |
| 166 | RIOK2 | 4.96 | 2.17 | 2.79 | 2.3 | 1.8 | -1.3 | 110 |

Continued on next page

**Table B.6 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 167 | ROS1 | 1.24 | 0. | 0. | 400. | 400. | -1. | 110 |
| 168 | RPS6KA1 | 6.51 | 3.41 | 2.79 | 1.9 | 2.3 | 1.2 | 110 |
| 169 | SCYL3 | 1.86 | 0.62 | 0.93 | 3. | 2. | -1.5 | 110 |
| 170 | SMG1 | 3.72 | 2.48 | 2.48 | 1.5 | 1.5 | -1. | 110 |
| 171 | STK19 | 0.93 | 0. | 0. | 300. | 300. | -1. | 110 |
| 172 | STK31 | 1.24 | 0.31 | 0. | 4. | 400. | 100. | 110 |
| 173 | STK32A | 1.55 | 0.62 | 0.31 | 2.5 | 5. | 2. | 110 |
| 174 | STK33 | 1.24 | 0. | 0. | 400. | 400. | -1. | 110 |
| 175 | STK38 | 8.99 | 3.72 | 4.34 | 2.4 | 2.1 | -1.2 | 110 |
| 176 | STK40 | 6.51 | 1.86 | 1.24 | 3.5 | 5.2 | 1.5 | 110 |
| 177 | STYK1 | 1.55 | 0.62 | 0.31 | 2.5 | 5. | 2. | 110 |
| 178 | TAOK2 | 2.17 | 0.62 | 1.24 | 3.5 | 1.8 | -2. | 110 |
| 179 | TRIB3 | 1.24 | 3.72 | 3.1 | -3. | -2.5 | 1.2 | 110 |
| 180 | TRPM7 | 2.48 | 0.93 | 1.24 | 2.7 | 2. | -1.3 | 110 |
| 181 | ULK3 | 4.96 | 1.86 | 2.48 | 2.7 | 2. | -1.3 | 110 |
| 182 | VRK1 | 6.2 | 4.34 | 3.72 | 1.4 | 1.7 | 1.2 | 110 |
| 183 | ACVR1 | 3.1 | 1.24 | 2.17 | 2.5 | 1.4 | -1.8 | 111 |
| 184 | ADCK2 | 5.27 | 1.55 | 2.48 | 3.4 | 2.1 | -1.6 | 111 |
| 185 | ADRBK1 | 8.68 | 4.03 | 12.71 | 2.2 | -1.5 | -3.2 | 111 |
| 186 | AKT2 | 8.06 | 5.58 | 2.79 | 1.4 | 2.9 | 2. | 111 |
| 187 | ARAF | 4.65 | 2.17 | 6.2 | 2.1 | -1.3 | -2.9 | 111 |
| 188 | BCKDK | 6.82 | 3.41 | 5.58 | 2. | 1.2 | -1.6 | 111 |
| 189 | BCR | 7.44 | 3.1 | 5.89 | 2.4 | 1.3 | -1.9 | 111 |
| 190 | BRAF | 7.44 | 2.17 | 3.41 | 3.4 | 2.2 | -1.6 | 111 |
| 191 | BRD2 | 5.58 | 2.79 | 4.03 | 2. | 1.4 | -1.4 | 111 |
| 192 | CAMK1D | 4.03 | 0. | 0.93 | 1300. | 4.3 | -300. | 111 |
| 193 | CAMK2G | 5.58 | 1.86 | 3.1 | 3. | 1.8 | -1.7 | 111 |
| 194 | CDC2 | 17.36 | 24.49 | 20.77 | -1.4 | -1.2 | 1.2 | 111 |
| 195 | CDC2L2 | 3.72 | 1.24 | 2.48 | 3. | 1.5 | -2. | 111 |
| 196 | CDK4 | 13.33 | 217.62 | 22.94 | -16.3 | -1.7 | 9.5 | 111 |
| 197 | CHUK | 4.03 | 2.48 | 7.13 | 1.6 | -1.8 | -2.9 | 111 |
| 198 | CIT | 1.24 | 2.48 | 0.31 | -2. | 4. | 8. | 111 |
| 199 | CPNE3 | 4.65 | 2.17 | 11.78 | 2.1 | -2.5 | -5.4 | 111 |
| 200 | CSNK1D | 33.79 | 21.39 | 12.4 | 1.6 | 2.7 | 1.7 | 111 |
| 201 | CSNK1E | 11.16 | 5.89 | 12.09 | 1.9 | -1.1 | -2.1 | 111 |
| 202 | DAPK1 | 4.03 | 2.17 | 0.93 | 1.9 | 4.3 | 2.3 | 111 |
| 203 | DAPK3 | 6.2 | 2.17 | 8.06 | 2.9 | -1.3 | -3.7 | 111 |
| 204 | DDR1 | 10.85 | 3.72 | 2.17 | 2.9 | 5. | 1.7 | 111 |
| 205 | EGFR | 123.69 | 124.62 | 4.65 | -1. | 26.6 | 26.8 | 111 |
| 206 | EIF2AK1 | 19.84 | 7.75 | 8.68 | 2.6 | 2.3 | -1.1 | 111 |
| 207 | EIF2AK2 | 15.81 | 5.89 | 10.23 | 2.7 | 1.5 | -1.7 | 111 |
| 208 | EIF2AK4 | 5.89 | 4.34 | 6.82 | 1.4 | -1.2 | -1.6 | 111 |

Continued on next page

**Table B.6 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 209 | EPHB4 | 7.13 | 0.93 | 3.1 | 7.7 | 2.3 | -3.3 | 111 |
| 210 | ERBB2 | 33.17 | 3.1 | 4.34 | 10.7 | 7.6 | -1.4 | 111 |
| 211 | ERBB3 | 18.6 | 7.44 | 5.27 | 2.5 | 3.5 | 1.4 | 111 |
| 212 | GRK6 | 4.34 | 5.27 | 3.41 | -1.2 | 1.3 | 1.5 | 111 |
| 213 | HIPK3 | 2.79 | 1.55 | 3.72 | 1.8 | -1.3 | -2.4 | 111 |
| 214 | HUS1 | 5.58 | 2.48 | 8.68 | 2.2 | -1.6 | -3.5 | 111 |
| 215 | ILK | 3.41 | 0.93 | 2.48 | 3.7 | 1.4 | -2.7 | 111 |
| 216 | IRAK1 | 7.44 | 12.4 | 13.64 | -1.7 | -1.8 | -1.1 | 111 |
| 217 | JAK1 | 3.72 | 5.89 | 2.79 | -1.6 | 1.3 | 2.1 | 111 |
| 218 | MAP2K1 | 8.68 | 10.54 | 13.33 | -1.2 | -1.5 | -1.3 | 111 |
| 219 | MAP2K2 | 11.16 | 9.92 | 26.04 | 1.1 | -2.3 | -2.6 | 111 |
| 220 | MAP2K4 | 0.93 | 3.1 | 4.34 | -3.3 | -4.7 | -1.4 | 111 |
| 221 | MAP3K1 | 4.03 | 2.17 | 0.93 | 1.9 | 4.3 | 2.3 | 111 |
| 222 | MAP3K7 | 4.96 | 2.79 | 4.03 | 1.8 | 1.2 | -1.4 | 111 |
| 223 | MAPK1 | 14.57 | 4.96 | 10.54 | 2.9 | 1.4 | -2.1 | 111 |
| 224 | MAPK13 | 53.63 | 4.34 | 6.2 | 12.4 | 8.6 | -1.4 | 111 |
| 225 | MAPK14 | 16.12 | 3.41 | 5.27 | 4.7 | 3.1 | -1.5 | 111 |
| 226 | MAPK3 | 11.78 | 3.41 | 5.89 | 3.5 | 2. | -1.7 | 111 |
| 227 | MAPK6 | 4.96 | 3.41 | 7.13 | 1.5 | -1.4 | -2.1 | 111 |
| 228 | MAPK8 | 2.48 | 1.55 | 4.03 | 1.6 | -1.6 | -2.6 | 111 |
| 229 | MELK | 11.78 | 6.51 | 7.75 | 1.8 | 1.5 | -1.2 | 111 |
| 230 | MERTK | 4.65 | 1.86 | 0.62 | 2.5 | 7.5 | 3. | 111 |
| 231 | MET | 15.81 | 62. | 25.73 | -3.9 | -1.6 | 2.4 | 111 |
| 232 | MKNK2 | 6.51 | 3.1 | 5.58 | 2.1 | 1.2 | -1.8 | 111 |
| 233 | MST1R | 9.61 | 3.1 | 6.2 | 3.1 | 1.5 | -2. | 111 |
| 234 | MST4 | 3.72 | 5.27 | 0. | -1.4 | 1200. | 1700. | 111 |
| 235 | MYLK | 0.31 | 4.03 | 19.53 | -13. | -63. | -4.8 | 111 |
| 236 | NEK7 | 6.51 | 4.34 | 5.58 | 1.5 | 1.2 | -1.3 | 111 |
| 237 | NEK9 | 3.41 | 1.24 | 2.17 | 2.8 | 1.6 | -1.8 | 111 |
| 238 | NRBP1 | 9.92 | 5.58 | 15.81 | 1.8 | -1.6 | -2.8 | 111 |
| 239 | PAK1 | 4.96 | 4.03 | 6.82 | 1.2 | -1.4 | -1.7 | 111 |
| 240 | PAK2 | 0.93 | 3.1 | 5.58 | -3.3 | -6. | -1.8 | 111 |
| 241 | PAK4 | 4.96 | 2.48 | 0.93 | 2. | 5.3 | 2.7 | 111 |
| 242 | PCTK1 | 5.89 | 4.65 | 10.85 | 1.3 | -1.8 | -2.3 | 111 |
| 243 | PDK1 | 1.24 | 4.34 | 3.1 | -3.5 | -2.5 | 1.4 | 111 |
| 244 | PDK2 | 4.65 | 0.93 | 2.48 | 5. | 1.9 | -2.7 | 111 |
| 245 | PIM2 | 4.96 | 0.62 | 1.86 | 8. | 2.7 | -3. | 111 |
| 246 | PINK1 | 2.48 | 1.24 | 5.27 | 2. | -2.1 | -4.2 | 111 |
| 247 | PKN2 | 4.03 | 5.89 | 2.48 | -1.5 | 1.6 | 2.4 | 111 |
| 248 | PLK1 | 1.86 | 4.96 | 6.51 | -2.7 | -3.5 | -1.3 | 111 |
| 249 | PRKACA | 14.57 | 6.82 | 15.81 | 2.1 | -1.1 | -2.3 | 111 |
| 250 | PRKCD | 7.44 | 3.41 | 4.65 | 2.2 | 1.6 | -1.4 | 111 |

Continued on next page

**Table B.6 – continued from previous page**

| SN | Gene | 1 | 2 | 3 | 1:2 | 1:3 | 2:3 | Sig |
|---|---|---|---|---|---|---|---|---|
| 251 | PRKCZ | 5.27 | 1.24 | 0. | 4.2 | 1700. | 400. | 111 |
| 252 | PRPF4B | 10.23 | 4.03 | 5.58 | 2.5 | 1.8 | -1.4 | 111 |
| 253 | PTK2 | 10.85 | 7.75 | 4.96 | 1.4 | 2.2 | 1.6 | 111 |
| 254 | RAF1 | 5.58 | 2.48 | 6.51 | 2.2 | -1.2 | -2.6 | 111 |
| 255 | RIOK3 | 7.44 | 1.86 | 2.79 | 4. | 2.7 | -1.5 | 111 |
| 256 | RIPK1 | 3.72 | 1.55 | 2.79 | 2.4 | 1.3 | -1.8 | 111 |
| 257 | RIPK2 | 2.48 | 8.37 | 9.92 | -3.4 | -4. | -1.2 | 111 |
| 258 | ROCK2 | 7.44 | 3.1 | 6.2 | 2.4 | 1.2 | -2. | 111 |
| 259 | RPS6KB2 | 3.41 | 1.86 | 5.58 | 1.8 | -1.6 | -3. | 111 |
| 260 | SCYL2 | 8.06 | 3.72 | 5.89 | 2.2 | 1.4 | -1.6 | 111 |
| 261 | SGK | 8.37 | 2.48 | 4.34 | 3.4 | 1.9 | -1.8 | 111 |
| 262 | SNF1LK | 4.65 | 1.24 | 3.41 | 3.8 | 1.4 | -2.8 | 111 |
| 263 | SRPK1 | 9.3 | 7.13 | 5.27 | 1.3 | 1.8 | 1.4 | 111 |
| 264 | SRPK2 | 10.23 | 3.41 | 7.44 | 3. | 1.4 | -2.2 | 111 |
| 265 | STK17A | 7.13 | 4.34 | 22.01 | 1.6 | -3.1 | -5.1 | 111 |
| 266 | STK17B | 13.95 | 3.1 | 4.65 | 4.5 | 3. | -1.5 | 111 |
| 267 | STK24 | 6.2 | 4.34 | 0. | 1.4 | 2000. | 1400. | 111 |
| 268 | STK38L | 4.65 | 2.48 | 3.41 | 1.9 | 1.4 | -1.4 | 111 |
| 269 | STK39 | 11.47 | 5.58 | 4.65 | 2.1 | 2.5 | 1.2 | 111 |
| 270 | TAOK1 | 8.68 | 2.17 | 4.65 | 4. | 1.9 | -2.1 | 111 |
| 271 | TAOK3 | 6.2 | 2.17 | 3.1 | 2.9 | 2. | -1.4 | 111 |
| 272 | TP53RK | 5.89 | 4.34 | 7.44 | 1.4 | -1.3 | -1.7 | 111 |
| 273 | TRIB1 | 7.13 | 1.55 | 3.72 | 4.6 | 1.9 | -2.4 | 111 |
| 274 | TRIB2 | 11.47 | 0. | 0.93 | 3700. | 12.3 | -300. | 111 |
| 275 | TRIO | 11.47 | 13.33 | 6.82 | -1.2 | 1.7 | 2. | 111 |
| 276 | UHMK1 | 6.82 | 5.89 | 11.47 | 1.2 | -1.7 | -1.9 | 111 |
| 277 | VRK3 | 4.34 | 0.31 | 1.55 | 14. | 2.8 | -5. | 111 |
| 278 | WNK1 | 16.12 | 5.89 | 7.44 | 2.7 | 2.2 | -1.3 | 111 |
| 279 | YES1 | 18.29 | 6.82 | 10.23 | 2.7 | 1.8 | -1.5 | 111 |

Table B.7: List of significantly changed gene in the group: (1 - $H322$, 2 - $H1703$), along with pairwise fold changes and significant change markers

| SN | Gene | 1 | 2 | 1:2 | Sig |
|---|---|---|---|---|---|
| 1 | ACVR1 | 0.31 | 1.55 | -5. | 1 |
| 2 | ACVR1B | 1.24 | 0.31 | 4. | 1 |
| 3 | ADCK2 | 1.55 | 3.41 | -2.2 | 1 |
| 4 | ADRBK1 | 9.92 | 4.96 | 2. | 1 |
| 5 | AKT2 | 5.27 | 4.03 | 1.3 | 1 |
| 6 | ALS2CR2 | 0.62 | 1.55 | -2.5 | 1 |
| 7 | ARAF | 3.72 | 2.17 | 1.7 | 1 |
| 8 | AURKA | 3.1 | 1.55 | 2. | 1 |
| 9 | AURKB | 17.36 | 14.57 | 1.2 | 1 |
| 10 | BCKDK | 7.44 | 4.65 | 1.6 | 1 |
| 11 | BCR | 3.41 | 1.86 | 1.8 | 1 |
| 12 | BMP2K | 1.86 | 0.62 | 3. | 1 |
| 13 | BMPR1A | 4.65 | 6.51 | -1.4 | 1 |
| 14 | BMPR1B | 1.24 | 0.31 | 4. | 1 |
| 15 | BMPR2 | 1.86 | 2.79 | -1.5 | 1 |
| 16 | BRAF | 1.86 | 3.41 | -1.8 | 1 |
| 17 | BRD2 | 7.44 | 4.96 | 1.5 | 1 |
| 18 | BUB1 | 6.51 | 2.79 | 2.3 | 1 |
| 19 | BUB1B | 11.16 | 4.03 | 2.8 | 1 |
| 20 | CAMK1 | 0. | 3.72 | -1200. | 1 |
| 21 | CAMK2D | 13.02 | 3.41 | 3.8 | 1 |
| 22 | CAMK2G | 4.34 | 2.48 | 1.8 | 1 |
| 23 | CASK | 1.55 | 0.62 | 2.5 | 1 |
| 24 | CDC2 | 53.32 | 26.04 | 2. | 1 |
| 25 | CDC2L5 | 4.03 | 3.1 | 1.3 | 1 |
| 26 | CDC2L6 | 0.93 | 3.72 | -4. | 1 |
| 27 | CDC42BPB | 3.41 | 2.48 | 1.4 | 1 |
| 28 | CDK10 | 5.89 | 2.48 | 2.4 | 1 |
| 29 | CDK4 | 17.05 | 13.95 | 1.2 | 1 |
| 30 | CDK5 | 0.31 | 2.17 | -7. | 1 |
| 31 | CDK6 | 7.44 | 1.24 | 6. | 1 |
| 32 | CDK7 | 0.93 | 3.1 | -3.3 | 1 |
| 33 | CDK8 | 3.1 | 4.03 | -1.3 | 1 |
| 34 | CDK9 | 1.86 | 2.79 | -1.5 | 1 |
| 35 | CHEK1 | 4.65 | 2.17 | 2.1 | 1 |
| 36 | CHUK | 3.1 | 6.2 | -2. | 1 |
| 37 | CIT | 0.93 | 1.86 | -2. | 1 |
| 38 | CLK2 | 3.41 | 1.55 | 2.2 | 1 |
| 39 | CLK4 | 0.62 | 1.86 | -3. | 1 |
| 40 | COL4A3BP | 2.17 | 3.1 | -1.4 | 1 |

Continued on next page

**Table B.7 – continued from previous page**

| SN | Gene | 1 | 2 | 1:2 | Sig |
|---|---|---|---|---|---|
| 41 | CPNE3 | 13.64 | 5.27 | 2.6 | 1 |
| 42 | CSK | 17.67 | 2.48 | 7.1 | 1 |
| 43 | CSNK1A1 | 4.65 | 5.58 | -1.2 | 1 |
| 44 | CSNK1D | 16.12 | 21.08 | -1.3 | 1 |
| 45 | CSNK1E | 5.27 | 10.54 | -2. | 1 |
| 46 | CSNK1G1 | 3.41 | 2.48 | 1.4 | 1 |
| 47 | CSNK1G3 | 3.41 | 5.27 | -1.5 | 1 |
| 48 | CSNK2A1 | 2.17 | 3.72 | -1.7 | 1 |
| 49 | DAPK3 | 1.55 | 3.72 | -2.4 | 1 |
| 50 | DDR1 | 2.79 | 0.62 | 4.5 | 1 |
| 51 | DYRK1A | 3.41 | 2.17 | 1.6 | 1 |
| 52 | DYRK2 | 3.72 | 1.24 | 3. | 1 |
| 53 | DYRK3 | 0. | 1.55 | -500. | 1 |
| 54 | EGFR | 4.65 | 2.48 | 1.9 | 1 |
| 55 | EIF2AK2 | 9.61 | 7.13 | 1.3 | 1 |
| 56 | EIF2AK3 | 2.48 | 0.31 | 8. | 1 |
| 57 | EIF2AK4 | 8.06 | 4.96 | 1.6 | 1 |
| 58 | EPHB4 | 0.31 | 1.24 | -4. | 1 |
| 59 | ERBB2 | 3.41 | 2.48 | 1.4 | 1 |
| 60 | ERBB3 | 5.58 | 0. | 1800. | 1 |
| 61 | FGFR1 | 0. | 5.58 | -1800. | 1 |
| 62 | FLJ13149 | 0.31 | 1.55 | -5. | 1 |
| 63 | FLJ23356 | 8.99 | 2.79 | 3.2 | 1 |
| 64 | FRK | 0.93 | 0. | 300. | 1 |
| 65 | FYN | 0.31 | 1.24 | -4. | 1 |
| 66 | GSK3B | 12.71 | 6.2 | 2. | 1 |
| 67 | HIPK3 | 9.3 | 1.24 | 7.5 | 1 |
| 68 | HUS1 | 7.13 | 4.65 | 1.5 | 1 |
| 69 | IGF1R | 4.65 | 0. | 1500. | 1 |
| 70 | INSR | 0.62 | 1.55 | -2.5 | 1 |
| 71 | IRAK1 | 11.47 | 8.37 | 1.4 | 1 |
| 72 | IRAK2 | 0.31 | 1.24 | -4. | 1 |
| 73 | JAK1 | 1.86 | 3.41 | -1.8 | 1 |
| 74 | KIAA0971 | 1.55 | 0.62 | 2.5 | 1 |
| 75 | KIAA1804 | 1.55 | 0.31 | 5. | 1 |
| 76 | LATS2 | 1.55 | 2.79 | -1.8 | 1 |
| 77 | LMTK2 | 8.06 | 2.48 | 3.2 | 1 |
| 78 | MAP2K1 | 36.58 | 7.75 | 4.7 | 1 |
| 79 | MAP2K2 | 25.42 | 20.46 | 1.2 | 1 |
| 80 | MAP2K3 | 5.58 | 2.17 | 2.6 | 1 |
| 81 | MAP2K4 | 4.34 | 2.48 | 1.8 | 1 |
| 82 | MAP2K7 | 1.86 | 0.93 | 2. | 1 |

Continued on next page

**Table B.7 – continued from previous page**

| SN | Gene | 1 | 2 | 1:2 | Sig |
|---|---|---|---|---|---|
| 83 | MAP3K1 | 1.55 | 0.31 | 5. | 1 |
| 84 | MAP3K12 | 0. | 1.55 | -500. | 1 |
| 85 | MAP3K3 | 0.62 | 2.48 | -4. | 1 |
| 86 | MAP3K5 | 1.55 | 0.31 | 5. | 1 |
| 87 | MAP4K2 | 2.48 | 1.24 | 2. | 1 |
| 88 | MAP4K3 | 4.34 | 1.55 | 2.8 | 1 |
| 89 | MAP4K4 | 7.75 | 3.72 | 2.1 | 1 |
| 90 | MAP4K5 | 7.75 | 4.03 | 1.9 | 1 |
| 91 | MAPK1 | 11.78 | 9.3 | 1.3 | 1 |
| 92 | MAPK11 | 0. | 0.93 | -300. | 1 |
| 93 | MAPK12 | 0.93 | 4.96 | -5.3 | 1 |
| 94 | MAPK13 | 14.88 | 0. | 4800. | 1 |
| 95 | MAPK3 | 7.44 | 4.03 | 1.8 | 1 |
| 96 | MAPK6 | 3.72 | 5.89 | -1.6 | 1 |
| 97 | MAPK8 | 3.41 | 5.27 | -1.5 | 1 |
| 98 | MAPK9 | 4.96 | 5.89 | -1.2 | 1 |
| 99 | MAPKAPK5 | 4.34 | 3.1 | 1.4 | 1 |
| 100 | MARK1 | 3.1 | 0.93 | 3.3 | 1 |
| 101 | MARK2 | 13.64 | 3.41 | 4. | 1 |
| 102 | MARK4 | 1.55 | 2.48 | -1.6 | 1 |
| 103 | MASTL | 4.65 | 2.17 | 2.1 | 1 |
| 104 | MELK | 11.47 | 6.51 | 1.8 | 1 |
| 105 | MERTK | 0.31 | 1.24 | -4. | 1 |
| 106 | MET | 4.96 | 1.55 | 3.2 | 1 |
| 107 | MGC5297 | 0.93 | 2.48 | -2.7 | 1 |
| 108 | MINK1 | 4.03 | 4.96 | -1.2 | 1 |
| 109 | MKNK1 | 3.72 | 1.86 | 2. | 1 |
| 110 | MKNK2 | 4.65 | 5.89 | -1.3 | 1 |
| 111 | MST1R | 10.23 | 0. | 3300. | 1 |
| 112 | MYLK | 0. | 1.24 | -400. | 1 |
| 113 | NEK11 | 0.31 | 2.17 | -7. | 1 |
| 114 | NEK2 | 4.03 | 2.17 | 1.9 | 1 |
| 115 | NEK3 | 0.62 | 3.1 | -5. | 1 |
| 116 | NEK4 | 1.24 | 0.31 | 4. | 1 |
| 117 | NEK7 | 13.02 | 6.2 | 2.1 | 1 |
| 118 | NLK | 0.62 | 1.86 | -3. | 1 |
| 119 | NRBP1 | 15.5 | 17.98 | -1.2 | 1 |
| 120 | NRBP2 | 0.31 | 1.86 | -6. | 1 |
| 121 | PAK2 | 5.27 | 8.06 | -1.5 | 1 |
| 122 | PAK4 | 3.72 | 2.48 | 1.5 | 1 |
| 123 | PBK | 6.82 | 3.41 | 2. | 1 |
| 124 | PCTK1 | 10.85 | 5.89 | 1.8 | 1 |

Continued on next page

**Table B.7 – continued from previous page**

| SN | Gene | 1 | 2 | 1:2 | Sig |
|---|---|---|---|---|---|
| 125 | PDGFRA | 0. | 203.05 | -65500. | 1 |
| 126 | PDIK1L | 2.17 | 0.93 | 2.3 | 1 |
| 127 | PDK1 | 1.86 | 0.93 | 2. | 1 |
| 128 | PDK3 | 1.24 | 0.31 | 4. | 1 |
| 129 | PIM1 | 3.72 | 1.86 | 2. | 1 |
| 130 | PINK1 | 0.31 | 1.24 | -4. | 1 |
| 131 | PKN2 | 5.58 | 3.72 | 1.5 | 1 |
| 132 | PKN3 | 0.62 | 1.55 | -2.5 | 1 |
| 133 | PLK1 | 12.71 | 6.2 | 2. | 1 |
| 134 | PLK2 | 3.41 | 15.5 | -4.5 | 1 |
| 135 | PRKAA1 | 10.54 | 13.95 | -1.3 | 1 |
| 136 | PRKAA2 | 2.48 | 1.24 | 2. | 1 |
| 137 | PRKACA | 10.23 | 19.22 | -1.9 | 1 |
| 138 | PRKACB | 7.13 | 2.48 | 2.9 | 1 |
| 139 | PRKCA | 0. | 1.24 | -400. | 1 |
| 140 | PRKCD | 3.1 | 1.24 | 2.5 | 1 |
| 141 | PRKCH | 7.44 | 0.93 | 8. | 1 |
| 142 | PRKD1 | 1.86 | 0.93 | 2. | 1 |
| 143 | PRKDC | 16.74 | 11.16 | 1.5 | 1 |
| 144 | PSKH1 | 2.17 | 0.62 | 3.5 | 1 |
| 145 | PTK2 | 30.07 | 9.3 | 3.2 | 1 |
| 146 | PTK6 | 1.24 | 0. | 400. | 1 |
| 147 | PTK7 | 2.79 | 0.62 | 4.5 | 1 |
| 148 | RAF1 | 6.51 | 8.37 | -1.3 | 1 |
| 149 | RIOK3 | 3.41 | 4.96 | -1.5 | 1 |
| 150 | RIPK1 | 2.17 | 3.72 | -1.7 | 1 |
| 151 | RIPK2 | 7.44 | 8.68 | -1.2 | 1 |
| 152 | RIPK4 | 5.89 | 0. | 1900. | 1 |
| 153 | ROCK1 | 7.13 | 11.16 | -1.6 | 1 |
| 154 | RPS6KA1 | 4.65 | 2.79 | 1.7 | 1 |
| 155 | RPS6KA4 | 6.2 | 2.17 | 2.9 | 1 |
| 156 | RPS6KA6 | 0.31 | 1.24 | -4. | 1 |
| 157 | RPS6KB1 | 2.17 | 3.41 | -1.6 | 1 |
| 158 | RPS6KB2 | 3.1 | 1.24 | 2.5 | 1 |
| 159 | SCYL1 | 4.65 | 2.48 | 1.9 | 1 |
| 160 | SGK | 1.24 | 109.74 | -88.5 | 1 |
| 161 | SGK3 | 1.86 | 0.31 | 6. | 1 |
| 162 | SMG1 | 3.72 | 5.27 | -1.4 | 1 |
| 163 | SNF1LK2 | 4.03 | 2.17 | 1.9 | 1 |
| 164 | SRC | 6.2 | 0.62 | 10. | 1 |
| 165 | SRPK2 | 1.86 | 8.06 | -4.3 | 1 |
| 166 | STK10 | 0. | 1.24 | -400. | 1 |

Continued on next page

**Table B.7 – continued from previous page**

| SN | Gene | 1 | 2 | 1:2 | Sig |
|---|---|---|---|---|---|
| 167 | STK11 | 1.55 | 3.1 | -2. | 1 |
| 168 | STK17A | 8.99 | 1.86 | 4.8 | 1 |
| 169 | STK17B | 4.34 | 1.24 | 3.5 | 1 |
| 170 | STK24 | 8.68 | 7.44 | 1.2 | 1 |
| 171 | STK3 | 3.1 | 1.55 | 2. | 1 |
| 172 | STK38 | 3.72 | 4.96 | -1.3 | 1 |
| 173 | STK39 | 6.51 | 3.41 | 1.9 | 1 |
| 174 | STK40 | 10.54 | 2.17 | 4.9 | 1 |
| 175 | TAF1 | 0.93 | 2.17 | -2.3 | 1 |
| 176 | TAF1L | 0.93 | 1.86 | -2. | 1 |
| 177 | TAOK1 | 5.27 | 7.13 | -1.4 | 1 |
| 178 | TESK1 | 0.62 | 1.55 | -2.5 | 1 |
| 179 | TGFBR2 | 11.47 | 24.49 | -2.1 | 1 |
| 180 | TLK2 | 1.86 | 3.1 | -1.7 | 1 |
| 181 | TP53RK | 4.96 | 3.41 | 1.5 | 1 |
| 182 | TRIB1 | 2.48 | 0. | 800. | 1 |
| 183 | TRIB2 | 5.89 | 0. | 1900. | 1 |
| 184 | TRIB3 | 0.93 | 6.2 | -6.7 | 1 |
| 185 | TRIO | 6.82 | 23.25 | -3.4 | 1 |
| 186 | TRPM7 | 0.93 | 1.86 | -2. | 1 |
| 187 | TTK | 12.71 | 5.89 | 2.2 | 1 |
| 188 | ULK3 | 8.99 | 2.17 | 4.1 | 1 |
| 189 | VRK1 | 8.06 | 2.48 | 3.2 | 1 |
| 190 | VRK2 | 3.41 | 1.55 | 2.2 | 1 |
| 191 | WNK1 | 13.95 | 20.15 | -1.4 | 1 |
| 192 | ZAK | 0.31 | 1.86 | -6. | 1 |

Table B.8: List of significantly changed gene in the group: (1 - $H827$, 2 - $H827Cripto$), along with pairwise fold changes and significant change markers

| SN | Gene | 1 | 2 | 1:2 | Sig |
|---|---|---|---|---|---|
| 1 | AAK1 | 2.79 | 3.72 | -1.3 | 1 |
| 2 | ACVR1 | 1.24 | 3.1 | -2.5 | 1 |
| 3 | ACVR1B | 0.62 | 2.17 | -3.5 | 1 |
| 4 | ADRBK1 | 4.03 | 6.2 | -1.5 | 1 |
| 5 | AKT2 | 5.58 | 8.06 | -1.4 | 1 |
| 6 | AKT3 | 0. | 2.79 | -900. | 1 |
| 7 | ARAF | 2.17 | 3.1 | -1.4 | 1 |
| 8 | AURKB | 15.5 | 4.03 | 3.8 | 1 |
| 9 | AXL | 0.93 | 13.02 | -14. | 1 |
| 10 | BCR | 3.1 | 4.03 | -1.3 | 1 |
| 11 | BMP2K | 1.55 | 2.48 | -1.6 | 1 |
| 12 | BMPR1A | 1.55 | 3.1 | -2. | 1 |
| 13 | BMPR2 | 2.17 | 6.2 | -2.9 | 1 |
| 14 | BRAF | 2.17 | 4.03 | -1.9 | 1 |
| 15 | BRD2 | 2.79 | 5.27 | -1.9 | 1 |
| 16 | BRDT | 1.55 | 2.79 | -1.8 | 1 |
| 17 | BUB1 | 5.58 | 2.17 | 2.6 | 1 |
| 18 | BUB1B | 5.89 | 3.1 | 1.9 | 1 |
| 19 | CAMK2D | 3.72 | 5.27 | -1.4 | 1 |
| 20 | CAMK2G | 1.86 | 3.72 | -2. | 1 |
| 21 | CAMKK2 | 2.17 | 3.72 | -1.7 | 1 |
| 22 | CDC2 | 24.49 | 13.02 | 1.9 | 1 |
| 23 | CDC2L5 | 2.48 | 3.72 | -1.5 | 1 |
| 24 | CDC2L6 | 1.55 | 2.79 | -1.8 | 1 |
| 25 | CDC42BPB | 2.48 | 4.65 | -1.9 | 1 |
| 26 | CDK10 | 6.82 | 13.33 | -2. | 1 |
| 27 | CDK4 | 217.62 | 295.74 | -1.4 | 1 |
| 28 | CDK6 | 3.1 | 4.65 | -1.5 | 1 |
| 29 | CDK8 | 1.24 | 3.1 | -2.5 | 1 |
| 30 | CDK9 | 2.17 | 3.41 | -1.6 | 1 |
| 31 | CDKL5 | 0.31 | 1.24 | -4. | 1 |
| 32 | CHUK | 2.48 | 3.72 | -1.5 | 1 |
| 33 | CIT | 2.48 | 1.24 | 2. | 1 |
| 34 | CPNE3 | 2.17 | 4.03 | -1.9 | 1 |
| 35 | CSNK1D | 21.39 | 33.79 | -1.6 | 1 |
| 36 | CSNK1E | 5.89 | 12.71 | -2.2 | 1 |
| 37 | CSNK1G1 | 0.93 | 2.48 | -2.7 | 1 |
| 38 | CSNK1G3 | 2.48 | 3.72 | -1.5 | 1 |
| 39 | CSNK2A1 | 3.41 | 5.58 | -1.6 | 1 |
| 40 | DAPK1 | 2.17 | 16.12 | -7.4 | 1 |

Continued on next page

**Table B.8 – continued from previous page**

| SN | Gene | 1 | 2 | 1:2 | Sig |
|---|---|---|---|---|---|
| 41 | DAPK3 | 2.17 | 4.65 | -2.1 | 1 |
| 42 | DDR1 | 3.72 | 5.89 | -1.6 | 1 |
| 43 | DYRK2 | 2.17 | 5.58 | -2.6 | 1 |
| 44 | EGFR | 124.62 | 160.89 | -1.3 | 1 |
| 45 | EIF2AK1 | 7.75 | 8.99 | -1.2 | 1 |
| 46 | EIF2AK2 | 5.89 | 8.99 | -1.5 | 1 |
| 47 | EPHA2 | 2.48 | 4.34 | -1.8 | 1 |
| 48 | EPHB2 | 0.31 | 1.24 | -4. | 1 |
| 49 | ERBB2 | 3.1 | 4.03 | -1.3 | 1 |
| 50 | ERBB3 | 7.44 | 24.18 | -3.2 | 1 |
| 51 | ERN1 | 0.93 | 2.17 | -2.3 | 1 |
| 52 | FGFR1 | 0. | 1.24 | -400. | 1 |
| 53 | FRK | 0.93 | 2.17 | -2.3 | 1 |
| 54 | FYN | 1.55 | 3.41 | -2.2 | 1 |
| 55 | GRK6 | 5.27 | 4.34 | 1.2 | 1 |
| 56 | GSG2 | 2.79 | 0.62 | 4.5 | 1 |
| 57 | GSK3A | 1.86 | 3.1 | -1.7 | 1 |
| 58 | GSK3B | 8.06 | 15.5 | -1.9 | 1 |
| 59 | HIPK1 | 0.93 | 1.86 | -2. | 1 |
| 60 | HIPK3 | 1.55 | 4.03 | -2.6 | 1 |
| 61 | HSPB8 | 0.62 | 1.86 | -3. | 1 |
| 62 | HUS1 | 2.48 | 4.65 | -1.9 | 1 |
| 63 | IGF1R | 0.62 | 1.55 | -2.5 | 1 |
| 64 | ILK | 0.93 | 1.86 | -2. | 1 |
| 65 | INSR | 0.93 | 2.79 | -3. | 1 |
| 66 | IRAK1 | 12.4 | 14.88 | -1.2 | 1 |
| 67 | JAK1 | 5.89 | 13.95 | -2.4 | 1 |
| 68 | LATS1 | 2.17 | 3.41 | -1.6 | 1 |
| 69 | LIMK1 | 8.06 | 12.71 | -1.6 | 1 |
| 70 | MAP2K1 | 10.54 | 17.67 | -1.7 | 1 |
| 71 | MAP2K2 | 9.92 | 13.33 | -1.3 | 1 |
| 72 | MAP2K4 | 3.1 | 2.17 | 1.4 | 1 |
| 73 | MAP3K1 | 2.17 | 3.41 | -1.6 | 1 |
| 74 | MAP3K13 | 0.62 | 1.86 | -3. | 1 |
| 75 | MAP3K2 | 2.79 | 6.82 | -2.4 | 1 |
| 76 | MAP3K5 | 0.31 | 2.79 | -9. | 1 |
| 77 | MAP3K7 | 2.79 | 4.34 | -1.6 | 1 |
| 78 | MAP4K3 | 2.79 | 5.58 | -2. | 1 |
| 79 | MAP4K4 | 11.78 | 15.5 | -1.3 | 1 |
| 80 | MAP4K5 | 8.06 | 13.64 | -1.7 | 1 |
| 81 | MAPK1 | 4.96 | 8.37 | -1.7 | 1 |
| 82 | MAPK10 | 0. | 0.93 | -300. | 1 |

Continued on next page

**Table B.8 – continued from previous page**

| SN | Gene | 1 | 2 | 1:2 | Sig |
|---|---|---|---|---|---|
| 83 | MAPK13 | 4.34 | 8.37 | -1.9 | 1 |
| 84 | MAPK14 | 3.41 | 5.89 | -1.7 | 1 |
| 85 | MAPK3 | 3.41 | 8.06 | -2.4 | 1 |
| 86 | MAPK6 | 3.41 | 6.82 | -2. | 1 |
| 87 | MAPK8 | 1.55 | 4.03 | -2.6 | 1 |
| 88 | MAPKAPK2 | 2.79 | 7.75 | -2.8 | 1 |
| 89 | MAPKAPK5 | 3.1 | 6.2 | -2. | 1 |
| 90 | MARK2 | 5.27 | 10.23 | -1.9 | 1 |
| 91 | MARK4 | 0.93 | 2.17 | -2.3 | 1 |
| 92 | MASTL | 2.79 | 1.86 | 1.5 | 1 |
| 93 | MELK | 6.51 | 3.72 | 1.8 | 1 |
| 94 | MERTK | 1.86 | 0.31 | 6. | 1 |
| 95 | MET | 62. | 75.02 | -1.2 | 1 |
| 96 | MGC16169 | 1.55 | 2.79 | -1.8 | 1 |
| 97 | MKNK1 | 2.79 | 4.03 | -1.4 | 1 |
| 98 | MKNK2 | 3.1 | 6.82 | -2.2 | 1 |
| 99 | MLKL | 2.17 | 3.1 | -1.4 | 1 |
| 100 | MST1R | 3.1 | 4.34 | -1.4 | 1 |
| 101 | MST4 | 5.27 | 7.44 | -1.4 | 1 |
| 102 | MYLK | 4.03 | 4.96 | -1.2 | 1 |
| 103 | NEK2 | 2.48 | 0.93 | 2.7 | 1 |
| 104 | NEK7 | 4.34 | 11.47 | -2.6 | 1 |
| 105 | NRBP1 | 5.58 | 9.92 | -1.8 | 1 |
| 106 | NUAK2 | 2.48 | 4.96 | -2. | 1 |
| 107 | OXSR1 | 1.55 | 2.79 | -1.8 | 1 |
| 108 | PAK1 | 4.03 | 8.06 | -2. | 1 |
| 109 | PAK2 | 3.1 | 5.89 | -1.9 | 1 |
| 110 | PBK | 4.03 | 1.86 | 2.2 | 1 |
| 111 | PCTK1 | 4.65 | 7.44 | -1.6 | 1 |
| 112 | PCTK2 | 1.55 | 3.1 | -2. | 1 |
| 113 | PDK2 | 0.93 | 2.17 | -2.3 | 1 |
| 114 | PDPK1 | 0.93 | 2.17 | -2.3 | 1 |
| 115 | PFTK1 | 0.31 | 3.1 | -10. | 1 |
| 116 | PIM1 | 3.72 | 6.2 | -1.7 | 1 |
| 117 | PIM3 | 0.93 | 2.17 | -2.3 | 1 |
| 118 | PINK1 | 1.24 | 2.17 | -1.8 | 1 |
| 119 | PKN2 | 5.89 | 7.75 | -1.3 | 1 |
| 120 | PLK1 | 4.96 | 2.17 | 2.3 | 1 |
| 121 | PLK3 | 0.31 | 1.55 | -5. | 1 |
| 122 | PRKAA1 | 5.58 | 14.57 | -2.6 | 1 |
| 123 | PRKAA2 | 1.86 | 3.41 | -1.8 | 1 |
| 124 | PRKACA | 6.82 | 11.16 | -1.6 | 1 |

Continued on next page

**Table B.8 – continued from previous page**

| SN | Gene | 1 | 2 | 1:2 | Sig |
|---|---|---|---|---|---|
| 125 | PRKACB | 1.86 | 3.72 | -2. | 1 |
| 126 | PRKCA | 1.55 | 3.41 | -2.2 | 1 |
| 127 | PRKCD | 3.41 | 7.75 | -2.3 | 1 |
| 128 | PRKCI | 5.58 | 11.16 | -2. | 1 |
| 129 | PRKCZ | 1.24 | 2.48 | -2. | 1 |
| 130 | PRPF4B | 4.03 | 5.58 | -1.4 | 1 |
| 131 | PTK2 | 7.75 | 11.47 | -1.5 | 1 |
| 132 | RIOK2 | 2.17 | 1.24 | 1.8 | 1 |
| 133 | RIPK1 | 1.55 | 2.48 | -1.6 | 1 |
| 134 | RIPK2 | 8.37 | 13.02 | -1.6 | 1 |
| 135 | RIPK4 | 3.41 | 4.34 | -1.3 | 1 |
| 136 | ROCK1 | 2.79 | 6.2 | -2.2 | 1 |
| 137 | ROCK2 | 3.1 | 5.27 | -1.7 | 1 |
| 138 | RPS6KA3 | 1.86 | 7.13 | -3.8 | 1 |
| 139 | RPS6KB1 | 1.24 | 2.48 | -2. | 1 |
| 140 | RPS6KC1 | 1.24 | 2.79 | -2.2 | 1 |
| 141 | RYK | 1.24 | 2.17 | -1.8 | 1 |
| 142 | SCYL1 | 2.79 | 3.72 | -1.3 | 1 |
| 143 | SCYL2 | 3.72 | 7.13 | -1.9 | 1 |
| 144 | SCYL3 | 0.62 | 1.55 | -2.5 | 1 |
| 145 | SGK | 2.48 | 6.82 | -2.8 | 1 |
| 146 | SGK3 | 1.55 | 0.62 | 2.5 | 1 |
| 147 | SLK | 2.48 | 4.65 | -1.9 | 1 |
| 148 | SMG1 | 2.48 | 3.41 | -1.4 | 1 |
| 149 | SNF1LK | 1.24 | 11.78 | -9.5 | 1 |
| 150 | SNF1LK2 | 1.55 | 3.1 | -2. | 1 |
| 151 | SRC | 2.79 | 5.89 | -2.1 | 1 |
| 152 | SRPK1 | 7.13 | 9.92 | -1.4 | 1 |
| 153 | SRPK2 | 3.41 | 5.27 | -1.5 | 1 |
| 154 | STK17A | 4.34 | 8.68 | -2. | 1 |
| 155 | STK17B | 3.1 | 4.96 | -1.6 | 1 |
| 156 | STK24 | 4.34 | 10.54 | -2.4 | 1 |
| 157 | STK25 | 2.79 | 3.72 | -1.3 | 1 |
| 158 | STK35 | 1.55 | 2.79 | -1.8 | 1 |
| 159 | STK38 | 3.72 | 7.44 | -2. | 1 |
| 160 | STK38L | 2.48 | 8.99 | -3.6 | 1 |
| 161 | STK39 | 5.58 | 13.64 | -2.4 | 1 |
| 162 | STK40 | 1.86 | 4.03 | -2.2 | 1 |
| 163 | STYK1 | 0.62 | 1.55 | -2.5 | 1 |
| 164 | TAOK1 | 2.17 | 4.34 | -2. | 1 |
| 165 | TAOK3 | 2.17 | 3.1 | -1.4 | 1 |
| 166 | TBK1 | 32.24 | 53.63 | -1.7 | 1 |

Continued on next page

**Table B.8 – continued from previous page**

| SN | Gene | 1 | 2 | 1:2 | Sig |
|---|---|---|---|---|---|
| 167 | TGFBR2 | 5.58 | 9.61 | -1.7 | 1 |
| 168 | TLK1 | 1.24 | 2.48 | -2. | 1 |
| 169 | TP53RK | 4.34 | 6.2 | -1.4 | 1 |
| 170 | TRIB1 | 1.55 | 8.06 | -5.2 | 1 |
| 171 | TRIB3 | 3.72 | 6.2 | -1.7 | 1 |
| 172 | TRIO | 13.33 | 17.67 | -1.3 | 1 |
| 173 | TTBK2 | 0.31 | 1.24 | -4. | 1 |
| 174 | TTK | 5.27 | 2.48 | 2.1 | 1 |
| 175 | UHMK1 | 5.89 | 8.68 | -1.5 | 1 |
| 176 | ULK1 | 1.24 | 4.03 | -3.2 | 1 |
| 177 | ULK3 | 1.86 | 3.1 | -1.7 | 1 |
| 178 | VRK1 | 4.34 | 2.17 | 2. | 1 |
| 179 | VRK2 | 1.55 | 2.79 | -1.8 | 1 |
| 180 | WNK1 | 5.89 | 11.47 | -1.9 | 1 |
| 181 | YES1 | 6.82 | 16.74 | -2.5 | 1 |

# Bibliography


[1] E.F. Keller. *Making sense of life: Explaining biological development with models, metaphors, and machines*. Harvard Univ Pr, 2003.

[2] Wikipedia. Circadian rhythm - wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Circadian_rhythm. [Online; accessed 31-March-2012].

[3] G.W. Litman, J.P. Cannon, and L.J. Dishaw. Reconstructing immune phylogeny: new perspectives. *Nature Reviews Immunology*, 5(11):866–879, 2005.

[4] G. Mayer. Immunology-chapter one: Innate (non-specific) immunity. *Microbiology and Immunology On-Line Textbook*, 2006.

[5] M. Berwick and P. Vineis. Markers of dna repair and susceptibility to cancer in humans: an epidemiologic review. *Journal of the National Cancer Institute*, 92(11):874–897, 2000.

[6] TS Chen and PS Chen. The myth of prometheus and the liver. *Journal of the Royal Society of Medicine*, 87(12):754, 1994.

[7] C. Power and J.E.J. Rasko. Whither prometheus' liver? greek myth and the science of regeneration. *Annals of internal medicine*, 149(6):421–426, 2008.

[8] K. Asonuma, J.C. Gilbert, J.E. Stein, T. Takeda, and J.P. Vacanti. Quantitation of transplanted hepatic mass necessary to cure the gunn rat model of hyperbilirubinemia. *Journal of pediatric surgery*, 27(3):298–301, 1992.

[9] G.K. Michalopoulos and M.C. DeFrances. Liver regeneration. *Science*, 276(5309):60–66, 1997.

[10] S. Mukherjee. *The emperor of all maladies: a biography of cancer*. Scribner Book Company, 2010.

[11] D Ferber. Immortality dies as bacteria show their age. *Science*, 307(5710):656–656, 2005.

[12] Z Nie, G Hu, G Wei, K Cui, A Yamane, W Resch, L Tessarollo, R Casellas, K Zhao, and D Levens. c-myc is a universal amplifier of gene expression. Submitted on 3-April-2012.

[13] I Wierstra and J Alves. The c-myc promoter: still mystery and challenge. *Adv Cancer Res*, 99:113–333, 2008.

[14] H.F. Judson. *The eighth day of creation: makers of the revolution in biology*. Cold Spring Harbor Laboratory Pr, 1996.

[15] R Dulbecco and M Vogt. Evidence for a ring structure of polyoma virus dna. *Proc Natl Acad Sci U S A*, 50:236–243, Aug 1963.

[16] J C Wang. Interaction between dna and an escherichia coli protein omega. *J Mol Biol*, 55(3):523–533, Feb 1971.

[17] A V Vologodskii, S D Levene, K V Klenin, M Frank-Kamenetskii, and N R Cozzarelli. Conformational and thermodynamic properties of supercoiled dna. *J Mol Biol*, 227(4):1224–1243, Oct 1992.

[18] A V Vologodskii and N R Cozzarelli. Conformational and thermodynamic properties of supercoiled dna. *Annu Rev Biophys Biomol Struct*, 23:609–643, 1994.

[19] C Lavelle. Forces and torques in the nucleus: chromatin under mechanical constraints. *Biochem Cell Biol*, 87(1):307–322, Feb 2009.

[20] Wikipedia. Hydrogen bond - wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Hydrogen_bond. [Online; accessed 31-March-2012].

[21] WIP Mainwaring. *Nucleic acid biochemistry and molecular biology*. Blackwell Scientific Publications Ltd, 1982.

[22] A.D. Bates and A. Maxwell. *DNA Topology*. Oxford University Press, USA, 2005.

[23] X Darzacq, Y Shav-Tal, V de Turris, Y Brody, S M Shenoy, R D Phair, and R H Singer. In vivo dynamics of rna polymerase ii transcription. *Nat Struct Mol Biol*, 14(9):796–806, Sep 2007.

[24] T R Strick, J F Allemand, D Bensimon, and V Croquette. Behavior of supercoiled dna. *Biophys J*, 74(4):2016–2028, Apr 1998.

[25] T R Strick, V Croquette, and D Bensimon. Homologous pairing in stretched supercoiled dna. *Proc Natl Acad Sci U S A*, 95(18):10579–10583, Sep 1998.

[26] J F Allemand, D Bensimon, and V Croquette. Stretching dna and rna to probe their interactions with proteins. *Curr Opin Struct Biol*, 13(3):266–274, Jun 2003.

[27] L.A. Johnston, D.A. Prober, B.A. Edgar, R.N. Eisenman, and P. Gallant. Drosophila myc regulates cellular growth during development. *Cell*, 98(6):779–790, 1999.

[28] F Kouzine, J Liu, S Sanford, H J Chung, and D Levens. The dynamic response of upstream dna to transcription-generated torsional stress. *Nat Struct Mol Biol*, 11(11):1092–1100, Nov 2004.

[29] J Liu, F Kouzine, Z Nie, H J Chung, Z Elisha-Feil, A Weber, K Zhao, and D Levens. The fuse/fbp/fir/tfiih system is a molecular machine programming a pulse of c-myc expression. *EMBO J*, 25(10):2119–2130, May 2006.

[30] F Kouzine and D Levens. Supercoil-driven dna structures regulate genetic transactions. *Front Biosci*, 12:4409–4423, 2007.

[31] F Kouzine, S Sanford, Z Elisha-Feil, and D Levens. The functional response of upstream dna to dynamic supercoiling in vivo. *Nat Struct Mol Biol*, 15(2):146–154, Feb 2008.

[32] Wikipedia. Encode - wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/ENCODE. [Online; accessed 31-March-2012].

[33] G. Church, D.W. Deamer, D. Branton, R. Baldarelli, and J. Kasianowicz. Characterization of individual polymer molecules based on monomer-interface interactions, August 18 1998. US Patent 5,795,782.

[34] C. Shaffer. Next-generation sequencing outpaces expectations. *Nature Biotechnology*, 25(2):149–149, 2007.

[35] B. McNally. *Next generation nanopore-based DNA sequencing*. PhD thesis, Boston University, 2012.

[36] V A Malkov, K A Serikawa, N Balantac, J Watters, G Geiss, A Mashadi-Hossein, and T Fare. Multiplexed measurements of gene signatures in different analytes using the nanostring ncounter assay system. *BMC Res Notes*, 2:80–80, 2009.

[37] Wikipedia. Kinase - wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Kinase. [Online; accessed 4-April-2012].

[38] M Muratani and W P Tansey. How the ubiquitin-proteasome system controls transcription. *Nat Rev Mol Cell Biol*, 4(3):192–201, Mar 2003.

[39] E Moreno. Is cell competition relevant to cancer? *Nat Rev Cancer*, 8(2):141–147, Feb 2008.

[40] L A Johnston. Competitive interactions between cells: death, growth, and geography. *Science*, 324(5935):1679–1682, Jun 2009.

[41] P Gallant. Myc, cell competition, and compensatory proliferation. *Cancer Res*, 65(15):6485–6487, Aug 2005.

[42] A Trumpp, Y Refaeli, T Oskarsson, S Gasser, M Murphy, G R Martin, and J M Bishop. c-myc regulates mammalian body size by controlling cell number but not cell size. *Nature*, 414(6865):768–773, Dec 2001.

[43] R.R. Sinden. Dna structure and function, 1994.

[44] G Elgar and T Vavouri. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet*, 24(7):344–352, Jul 2008.

[45] Wikipedia. Linking number - wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Linking_number. [Online; accessed 4-April-2012].

[46] R D Wells. Non-b dna conformations, mutagenesis and disease. *Trends Biochem Sci*, 32(6):271–278, Jun 2007.

[47] V.A. Bloomfield, D.M. Crothers, and I. Tinoco. *Nucleic acids: structures, properties, and functions*. Univ Science Books, 2000.

[48] L F Liu and J C Wang. Supercoiling of the dna template during transcription. *Proc Natl Acad Sci U S A*, 84(20):7024–7027, Oct 1987.

[49] B Gutiérrez-Medina, J O Andreasson, W J Greenleaf, A Laporta, and S M Block. An optical apparatus for rotation and trapping. *Methods Enzymol*, 475:377–404, 2010.

[50] P Nelson. Transport of torsional stress in dna. *Proc Natl Acad Sci U S A*, 96(25):14342–14347, Dec 1999.

[51] A L Gnatt, P Cramer, J Fu, D A Bushnell, and R D Kornberg. Structural basis of transcription: an rna polymerase ii elongation complex at 3.3 a resolution. *Science*, 292(5523):1876–1882, Jun 2001.

[52] M Le Bret. Monte carlo computation of the supercoiling energy, the sedimentation constant, and the radius of gyration of unknotted and knotted circular dna. *Biopolymers*, 19(3):619–637, Mar 1980.

[53] A V Vologodskii, V V Anshelevich, A V Lukashin, and M D Frank-Kamenetskii. Statistical mechanics of supercoils and the torsional stiffness of the dna double helix. *Nature*, 280(5720):294–298, Jul 1979.

[54] A V Vologodskii. Distributions of topological states in circular dna. *Mol Biol (Mosk)*, 35(2):285–297, Mar-Apr 2001.

[55] L Baranello, D Levens, A Gupta, and F Kouzine. The importance of being supercoiled: How dna mechanics regulate dynamic processes. *Biochim Biophys Acta*, Jan 2012.

[56] G W Hatfield and C J Benham. Dna topology-mediated control of global gene expression in escherichia coli. *Annu Rev Genet*, 36:175–203, 2002.

[57] A I Alexandrov, N R Cozzarelli, V F Holmes, A B Khodursky, B J Peter, L Postow, V Rybenkov, and A V Vologodskii. Mechanisms of separation of the complementary strands of dna during replication. *Genetica*, 106(1-2):131–140, 1999.

[58] N J Crisona, R Kanaar, T N Gonzalez, E L Zechiedrich, A Klippel, and N R Cozzarelli. Processive recombination by wild-type gin and an enhancer-independent mutant. insight into the mechanisms of recombination selectivity and strand exchange. *J Mol Biol*, 243(3):437–457, Oct 1994.

[59] E L Zechiedrich, A B Khodursky, S Bachellier, R Schneider, D Chen, D M Lilley, and N R Cozzarelli. Roles of topoisomerases in maintaining steady-state dna supercoiling in escherichia coli. *J Biol Chem*, 275(11):8103–8113, Mar 2000.

[60] A Travers and G Muskhelishvili. A common topology for bacterial and eukaryotic transcription initiation? *EMBO Rep*, 8(2):147–151, Feb 2007.

[61] A Worcel, S Strogatz, and D Riley. Structure of chromatin and the linking number of dna. *Proc Natl Acad Sci U S A*, 78(3):1461–1465, Mar 1981.

[62] K Luger, A W Mäder, R K Richmond, D F Sargent, and T J Richmond. Crystal structure of the nucleosome core particle at 2.8 a resolution. *Nature*, 389(6648):251–260, Sep 1997.

[63] C L Woodcock, A I Skoultchi, and Y Fan. Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res*, 14(1):17–25, 2006.

[64] J S Godde and J Widom. Chromatin structure of schizosaccharomyces pombe. a nucleosome repeat length that is shorter than the chromatosomal dna length. *J Mol Biol*, 226(4):1009–1025, Aug 1992.

[65] A Prunell. A topological approach to nucleosome structure and dynamics: the linking number paradox and other issues. *Biophys J*, 74(5):2531–2544, May 1998.

[66] A Bancaud, N Conde e Silva, M Barbi, G Wagner, J F Allemand, J Mozziconacci, C Lavelle, V Croquette, J M Victor, A Prunell, and J L Viovy. Structural plasticity of single chromatin fibers revealed by torsional manipulation. *Nat Struct Mol Biol*, 13(5):444–450, May 2006.

[67] L A Freeman and W T Garrard. Dna supercoiling in chromatin structure and gene expression. *Crit Rev Eukaryot Gene Expr*, 2(2):165–209, 1992.

[68] H Y Wu, S H Shyy, J C Wang, and L F Liu. Transcription generates positively and negatively supercoiled domains in the template. *Cell*, 53(3):433–440, May 1988.

[69] J C Wang. Cellular roles of dna topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol*, 3(6):430–440, Jun 2002.

[70] P Dröge. Protein tracking-induced supercoiling of dna: a tool to regulate dna transactions in vivo? *Bioessays*, 16(2):91–99, Feb 1994.

[71] C S Osborne, L Chakalova, K E Brown, D Carter, A Horton, E Debrand, B Goyenechea, J A Mitchell, S Lopes, W Reik, and P Fraser. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*, 36(10):1065–1071, Oct 2004.

[72] F J Iborra, D A Jackson, and P R Cook. Coupled transcription and translation within nuclei of mammalian cells. *Science*, 293(5532):1139–1142, Aug 2001.

[73] H Boeger, D A Bushnell, R Davis, J Griesenbeck, Y Lorch, J S Strattan, K D Westover, and R D Kornberg. Structural basis of eukaryotic gene transcription. *FEBS Lett*, 579(4):899–903, Feb 2005.

[74] P.R. Cook. A chromomeric model for nuclear and chromosome structure. *Journal of cell science*, 108(9):2927–2935, 1995.

[75] J Roca. Transcriptional inhibition by dna torsional stress. *Transcription*, 2(2):82–85, 3 2011.

[76] R S Joshi, B Piña, and J Roca. Positional dependence of transcriptional inhibition by dna torsional stress in yeast chromosomes. *EMBO J*, 29(4):740–748, Feb 2010.

[77] G Felsenfeld, D Clark, and V Studitsky. Transcription through nucleosomes. *Biophys Chem*, 86(2-3):231–237, Aug 2000.

[78] G D Bowman. Mechanisms of atp-dependent nucleosome sliding. *Curr Opin Struct Biol*, 20(1):73–81, Feb 2010.

[79] K Havas, A Flaus, M Phelan, R Kingston, P A Wade, D M Lilley, and T Owen-Hughes. Generation of superhelical torsion by atp-dependent chromatin remodeling activities. *Cell*, 103(7):1133–1142, Dec 2000.

[80] H A Cole, B H Howard, and D J Clark. Activation-induced disruption of nucleosome position clusters on the coding regions of gcn4-dependent genes extends into neighbouring genes. *Nucleic Acids Res*, 39(22):9521–9535, Dec 2011.

[81] Y Kim and D J Clark. Swi/snf-dependent long-range remodeling of yeast his3 chromatin. *Proc Natl Acad Sci U S A*, 99(24):15381–15386, Nov 2002.

[82] V V Philimonenko, J Zhao, S Iben, H Dingová, K Kyselá, M Kahle, H Zentgraf, W A Hofmann, P de Lanerolle, P Hozák, and I Grummt. Nuclear actin and myosin i are required for rna polymerase i transcription. *Nat Cell Biol*, 6(12):1165–1172, Dec 2004.

[83] P de Lanerolle, T Johnson, and W A Hofmann. Actin and myosin i in the nucleus: what next? *Nat Struct Mol Biol*, 12(9):742–746, Sep 2005.

[84] J Gore, Z Bryant, M Nöllmann, M U Le, N R Cozzarelli, and C Bustamante. Dna overwinds when stretched. *Nature*, 442(7104):836–839, Aug 2006.

[85] Y Brody, N Neufeld, N Bieberstein, S Z Causse, E M Böhnlein, K M Neugebauer, X Darzacq, and Y Shav-Tal. The in vivo kinetics of rna polymerase ii elongation during co-transcriptional splicing. *PLoS Biol*, 9(1), 2011.

[86] J Singh and R A Padgett. Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol*, 16(11):1128–1133, Nov 2009.

[87] A C Seila, J M Calabrese, S S Levine, G W Yeo, P B Rahl, R A Flynn, R A Young, and P A Sharp. Divergent transcription from active promoters. *Science*, 322(5909):1849–1851, Dec 2008.

[88] M R Gartenberg and J C Wang. Identification of barriers to rotation of dna segments in yeast from the topology of dna rings excised by an inducible site-specific recombinase. *Proc Natl Acad Sci U S A*, 90(22):10514–10518, Nov 1993.

[89] C Lavelle. Dna torsional stress propagates through chromatin fiber and participates in transcriptional regulation. *Nat Struct Mol Biol*, 15(2):123–125, Feb 2008.

[90] P Recouvreux, C Lavelle, M Barbi, N Conde E Silva, E Le Cam, J M Victor, and J L Viovy. Linker histones incorporation maintains chromatin fiber plasticity. *Biophys J*, 100(11):2726–2735, Jun 2011.

[91] C Lavelle, P Recouvreux, H Wong, A Bancaud, J L Viovy, A Prunell, and J M Victor. Right-handed nucleosome: myth or reality? *Cell*, 139(7):1216–1217, Dec 2009.

[92] J J Champoux. Dna topoisomerases: structure, function, and mechanism. *Annu Rev Biochem*, 70:369–413, 2001.

[93] S J Brill and R Sternglanz. Transcription-dependent dna supercoiling in yeast dna topoisomerase mutants. *Cell*, 54(3):403–411, Jul 1988.

[94] S L French, M L Sikes, R D Hontz, Y N Osheim, T E Lambert, A El Hage, M M Smith, D Tollervey, J S Smith, and A L Beyer. Distinguishing the roles of topoisomerases i and ii in relief of transcription-induced torsional stress in yeast rrna genes. *Mol Cell Biol*, 31(3):482–494, Feb 2011.

[95] A S Sperling, K S Jeong, T Kitada, and M Grunstein. Topoisomerase ii binds nucleosome-free dna and acts redundantly with topoisomerase i to enhance recruitment of rna pol ii in budding yeast. *Proc Natl Acad Sci U S A*, 108(31):12693–12698, Aug 2011.

[96] J Salceda, X Fernández, and J Roca. Topoisomerase ii, not topoisomerase i, is the proficient relaxase of nucleosomal dna. *EMBO J*, 25(11):2575–2583, Jun 2006.

[97] D A Koster, V Croquette, C Dekker, S Shuman, and N H Dekker. Friction and torque govern the relaxation of dna supercoils by eukaryotic topoisomerase ib. *Nature*, 434(7033):671–674, Mar 2005.

[98] J Vinograd, J Lebowitz, R Radloff, R Watson, and P Laipis. The twisted circular form of polyoma viral dna. *Proc Natl Acad Sci U S A*, 53(5):1104–1111, May 1965.

[99] H B Gray, W B Upholt, and J Vinograd. A buoyant method for the determination of the superhelix density of closed circular dna. *J Mol Biol*, 62(1):1–19, Nov 1971.

[100] J C Wang. The degree of unwinding of the dna helix by ethidium. i. titration of twisted pm2 dna molecules in alkaline cesium chloride density gradients. *J Mol Biol*, 89(4):783–801, Nov 1974.

[101] T R Strick, J F Allemand, D Bensimon, A Bensimon, and V Croquette. The elasticity of a single supercoiled dna molecule. *Science*, 271(5257):1835–1837, Mar 1996.

[102] D A Koster, A Crut, S Shuman, M A Bjornsti, and N H Dekker. Cellular strategies for regulating dna supercoiling: a single-molecule perspective. *Cell*, 142(4):519–530, Aug 2010.

[103] R R Sinden, J O Carlson, and D E Pettijohn. Torsional tension in the dna double helix measured with trimethylpsoralen in living escherichia coli cells: analogous measurements in insect and human cells. *Cell*, 21(3):773–783, Oct 1980.

[104] R R Sinden and D W Ussery. Analysis of dna structure in vivo using psoralen photobinding: measurement of supercoiling, topological domains, and dna-protein interactions. *Methods Enzymol*, 212:319–335, 1992.

[105] P R Kramer and R R Sinden. Measurement of unrestrained negative supercoiling and topological domain size in living human cells. *Biochemistry*, 36(11):3151–3158, Mar 1997.

[106] P R Kramer, O Bat, and R R Sinden. Measurement of localized dna supercoiling and topological domain size in eukaryotic cells. *Methods Enzymol*, 304:639–650, 1999.

[107] K Matsumoto and S Hirose. Visualization of unconstrained negative supercoils of dna on polytene chromosomes of drosophila. *J Cell Sci*, 117(Pt 17):3797–3805, Aug 2004.

[108] I Bermúdez, J García-Martínez, J E Pérez-Ortín, and J Roca. A method for genome-wide analysis of dna helical tension by means of psoralen-dna photobinding. *Nucleic Acids Res*, 38(19), Oct 2010.

[109] L Postow, C D Hardy, J Arsuaga, and N R Cozzarelli. Topological domain structure of the escherichia coli chromosome. *Genes Dev*, 18(14):1766–1779, Jul 2004.

[110] M R Gartenberg and J C Wang. Positive supercoiling of dna greatly diminishes mrna synthesis in yeast. *Proc Natl Acad Sci U S A*, 89(23):11461–11465, Dec 1992.

[111] L Baranello, D Bertozzi, M V Fogli, Y Pommier, and G Capranico. Dna topoisomerase i inhibition by camptothecin induces escape of rna polymerase ii from promoter-proximal pause site, antisense transcription and histone acetylation at the human hif-1alpha gene locus. *Nucleic Acids Res*, 38(1):159–171, Jan 2010.

[112] G Capranico, J Marinello, and L Baranello. Dissecting the transcriptional functions of human dna topoisomerase i by selective inhibitors: implications for physiological and therapeutic modulation of enzyme activity. *Biochim Biophys Acta*, 1806(2):240–250, Dec 2010.

[113] S J Petesch and J T Lis. Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at hsp70 loci. *Cell*, 134(1):74–84, Jul 2008.

[114] J Zlatanova and J M Victor. How are nucleosomes disrupted during transcription elongation? *HFSP J*, 3(6):373–378, Dec 2009.

[115] B Villeponteau, M Lundell, and H Martinson. Torsional stress promotes the dnaase i sensitivity of active genes. *Cell*, 39(3 Pt 2):469–478, Dec 1984.

[116] V.A. Bloomfield, D.M. Crothers, and I. Tinoco. *Physical chemistry of nucleic acids*. Harper & Row New York, 1974.

[117] T A Brooks, S Kendrick, and L Hurley. Making sense of g-quadruplex and i-motif functions in oncogene promoters. *FEBS J*, 277(17):3459–3469, Sep 2010.

[118] D Zhabinskaya and C J Benham. Theoretical analysis of the stress induced b-z transition in superhelical dna. *PLoS Comput Biol*, 7(1), 2011.

[119] J L Huppert and S Balasubramanian. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*, 35(2):406–413, 2007.

[120] A Siddiqui-Jain, C L Grand, D J Bearss, and L H Hurley. Direct evidence for a g-quadruplex in a promoter region and its targeting with a small molecule to repress c-myc transcription. *Proc Natl Acad Sci U S A*, 99(18):11593–11598, Sep 2002.

[121] T A Brooks and L H Hurley. The role of supercoiling in transcriptional control of myc and its importance in molecular therapeutics. *Nat Rev Cancer*, 9(12):849–861, Dec 2009.

[122] G A Michelotti, E F Michelotti, A Pullner, R C Duncan, D Eick, and D Levens. Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. *Mol Cell Biol*, 16(6):2656–2669, Jun 1996.

[123] Y Kohwi and T Kohwi-Shigematsu. Altered gene expression correlates with dna structure. *Genes Dev*, 5(12B):2547–2554, Dec 1991.

[124] T Tomonaga and D Levens. Activating transcription from single stranded dna. *Proc Natl Acad Sci U S A*, 93(12):5830–5835, Jun 1996.

[125] E F Michelotti, T Tomonaga, H Krutzsch, and D Levens. Cellular nucleic acid binding protein regulates the ct element of the human c-myc protooncogene. *J Biol Chem*, 270(16):9494–9499, Apr 1995.

[126] D Sun and L H Hurley. The importance of negative superhelicity in inducing the formation of g-quadruplex and i-motif structures in the c-myc promoter: implications for drug targeting and control of gene expression. *J Med Chem*, 52(9):2863–2874, May 2009.

[127] B Wittig, S Wölfl, T Dorbic, W Vahrson, and A Rich. Transcription of human c-myc in permeabilized nuclei is associated with formation of z-dna in three discrete regions of the gene. *EMBO J*, 11(12):4653–4663, Dec 1992.

[128] T Schwartz, J Behlke, K Lowenhaupt, U Heinemann, and A Rich. Structure of the dlm-1-z-dna complex reveals a conserved family of z-dna-binding proteins. *Nat Struct Biol*, 8(9):761–765, Sep 2001.

[129] H Liu, N Mulholland, H Fu, and K Zhao. Cooperative activity of brg1 and z-dna formation in chromatin remodeling. *Mol Cell Biol*, 26(7):2550–2559, Apr 2006.

[130] B Wong, S Chen, J A Kwon, and A Rich. Characterization of z-dna as a nucleosome-boundary element in yeast saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, 104(7):2229–2234, Feb 2007.

[131] R Liu, H Liu, X Chen, M Kirby, P O Brown, and K Zhao. Regulation of csf1 promoter by the swi/snf-like baf complex. *Cell*, 106(3):309–318, Aug 2001.

[132] J M Vilar and L Saiz. Dna looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev*, 15(2):136–144, Apr 2005.

[133] T Tomonaga, G A Michelotti, D Libutti, A Uy, B Sauer, and D Levens. Unrestraining genetic processes with a protein-dna hinge. *Mol Cell*, 1(5):759–764, Apr 1998.

[134] Y S Polikanov, V A Bondarenko, V Tchernaenko, Y I Jiang, L C Lutter, A Vologodskii, and V M Studitsky. Probability of the site juxtaposition determines the rate of protein-mediated dna looping. *Biophys J*, 93(8):2726–2731, Oct 2007.

[135] D Hanahan and R A Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, Mar 2011.

[136] E Segal, Y Fondufe-Mittendorf, L Chen, A Thåström, Y Field, I K Moore, J P Wang, and J Widom. A genomic code for nucleosome positioning. *Nature*, 442(7104):772–778, Aug 2006.

[137] E Segal, T Raveh-Sadka, M Schroeder, U Unnerstall, and U Gaul. Predicting expression patterns from regulatory sequence in drosophila segmentation. *Nature*, 451(7178):535–540, Jan 2008.

[138] A Gupta, F Kouzine, B Baranello, K Ben-Aissa, and D Levens. Dynamic supercoiling is differentially tuned by topoisomerases i and ii across the genome. Submitted on 18-April-2012.

[139] J Roca. The torsional state of dna within the chromosome. *Chromosoma*, 120(4):323–334, Aug 2011.

[140] L R Benjamin, H J Chung, S Sanford, F Kouzine, J Liu, and D Levens. Hierarchical mechanisms build the dna-binding specificity of fuse binding protein. *Proc Natl Acad Sci U S A*, 105(47):18296–18301, Nov 2008.

[141] Mickal Durand-Dubief, Jenna Persson, Ulrika Norman, Edgar Hartsuiker, and Karl Ekwall. Topoisomerase i regulates open chromatin and controls gene expression in vivo. *The EMBO Journal*, 29(13):2126–2134, July 2010. PMID: 20526281.

[142] A C Seila, L J Core, J T Lis, and P A Sharp. Divergent transcription: a new feature of active promoters. *Cell Cycle*, 8(16):2557–2564, Aug 2009.

[143] M Geertz, A Travers, S Mehandziska, P Sobetzko, S Chandra-Janga, N Shimamoto, and G Muskhelishvili. Structural coupling between rna polymerase composition and dna supercoiling in coordinating transcription: a global role for the omega subunit? *MBio*, 2(4), 2011.

[144] M Schwikardi and P Dröge. Site-specific recombination in mammalian cells catalyzed by gammadelta resolvase mutants: implications for the topology of episomal dna. *FEBS Lett*, 471(2-3):147–150, Apr 2000.

[145] R R Sinden, O Bat, and P R Kramer. Psoralen cross-linking as probe of torsional tension and topological domain size in vivo. *Methods*, 17(2):112–124, Feb 1999.

[146] K Takase, M Sawai, K Yamamoto, J Yata, Y Takasaki, H Teraoka, and K Tsukada. Reversible g1 arrest induced by dimethyl sulfoxide in human lymphoid cell lines: kinetics of the arrest and expression of the cell cycle marker proliferating cell nuclear antigen in raji cells. *Cell Growth Differ*, 3(8):515–521, Aug 1992.

[147] ENCODE Project Consortium, R M Myers, J Stamatoyannopoulos, M Snyder, I Dunham, R C Hardison, B E Bernstein, T R Gingeras, W J Kent, E Birney, B Wold, and G E Crawford. A user's guide to the encyclopedia of dna elements (encode). *PLoS Biol*, 9(4), Apr 2011.

[148] M K Raghuraman, E A Winzeler, D Collingwood, S Hunt, L Wodicka, A Conway, D J Lockhart, R W Davis, B J Brewer, and W L Fangman. Replication dynamics of the yeast genome. *Science*, 294(5540):115–121, Oct 2001.

[149] S M Vos, E M Tretter, B H Schmidt, and J M Berger. All tangled up: how cells direct, manage and exploit topoisomerase function. *Nat Rev Mol Cell Biol*, 12(12):827–841, Dec 2011.

[150] C J Li, L Averboukh, and A B Pardee. beta-lapachone, a novel dna topoisomerase i inhibitor with a mode of action different from camptothecin. *J Biol Chem*, 268(30):22463–22468, Oct 1993.

[151] B Frydman, L J Marton, J S Sun, K Neder, D T Witiak, A A Liu, H M Wang, Y Mao, H Y Wu, M M Sanders, and L F Liu. Induction of dna topoisomerase ii-mediated dna cleavage by beta-lapachone and related naphthoquinones. *Cancer Res*, 57(4):620–627, Feb 1997.

[152] Y Pommier. Topoisomerase i inhibitors: camptothecins and beyond. *Nat Rev Cancer*, 6(10):789–802, Oct 2006.

[153] B L Staker, K Hjerrild, M D Feese, C A Behnke, A B Burgin, and L Stewart. The mechanism of topoisomerase i poisoning by a camptothecin analog. *Proc Natl Acad Sci U S A*, 99(24):15387–15392, Nov 2002.

[154] D A Koster, K Palle, E S Bot, M A Bjornsti, and N H Dekker. Antitumour drugs impede dna uncoiling by topoisomerase i. *Nature*, 448(7150):213–217, Jul 2007.

[155] A Khobta, F Ferri, L Lotito, A Montecucco, R Rossi, and G Capranico. Early effects of topoisomerase i inhibition on rna polymerase ii along transcribed genes in human cells. *J Mol Biol*, 357(1):127–138, Mar 2006.

[156] J E Deweese and N Osheroff. The dna cleavage reaction of topoisomerase ii: wolf in sheep's clothing. *Nucleic Acids Res*, 37(3):738–748, Feb 2009.

[157] K D Bromberg, A B Burgin, and N Osheroff. A two-drug model for etoposide action against human topoisomerase iialpha. *J Biol Chem*, 278(9):7406–7412, Feb 2003.

[158] Y Timsit and P Várnai. Helical chirality: a link between local interactions and global topology in dna. *PLoS One*, 5(2), 2010.

[159] Y L Lyu, C P Lin, A M Azarova, L Cai, J C Wang, and L F Liu. Role of topoisomerase iibeta in the expression of developmentally regulated genes. *Mol Cell Biol*, 26(21):7929–7941, Nov 2006.

[160] J Roca and J C Wang. The probabilities of supercoil removal and decatenation by yeast dna topoisomerase ii. *Genes Cells*, 1(1):17–27, Jan 1996.

[161] H Wada and R R Netz. Plectoneme creation reduces the rotational friction of a polymer. *EPL (Europhysics Letters)*, 87:38001, 2009.

[162] B Treutlein, A Muschielok, J Andrecka, A Jawhari, C Buchen, D Kostrewa, F Hög, P Cramer, and J Michaelis. Dynamic architecture of a minimal rna polymerase ii open promoter complex. *Mol Cell*, Mar 2012.

[163] A C Cheung, S Sainsbury, and P Cramer. Structural basis of initial rna polymerase ii transcription. *EMBO J*, 30(23):4755–4763, Nov 2011.

[164] J Roca. The path of the dna along the dimer interface of topoisomerase ii. *J Biol Chem*, 279(24):25783–25788, Jun 2004.

[165] A Lesne, C Bécavin, and J M Victor. The condensed chromatin fiber: an allosteric chemo-mechanical machine for signal transduction and genome processing. *Phys Biol*, 9(1):013001–013001, Feb 2012.

[166] N J Fuda, M B Ardehali, and J T Lis. Defining mechanisms that regulate rna polymerase ii transcription in vivo. *Nature*, 461(7261):186–192, Sep 2009.

[167] S Unniraman and V Nagaraja. Axial distortion as a sensor of supercoil changes: a molecular model for the homeostatic regulation of dna gyrase. *J Genet*, 80(3):119–124, Dec 2001.

[168] A Revyakin, R H Ebright, and T R Strick. Promoter unwinding and promoter clearance by rna polymerase: detection by single-molecule dna nanomanipulation. *Proc Natl Acad Sci U S A*, 101(14):4776–4780, Apr 2004.

[169] J D Parvin and P A Sharp. Dna topology and a minimal set of basal factors for transcription by rna polymerase ii. *Cell*, 73(3):533–540, May 1993.

[170] B J Peter, J Arsuaga, A M Breier, A B Khodursky, P O Brown, and N R Cozzarelli. Genomic transcriptional response to loss of chromosomal supercoiling in escherichia coli. *Genome Biol*, 5(11), 2004.

[171] A Barski, S Cuddapah, K Cui, T Y Roh, D E Schones, Z Wang, G Wei, I Chepelev, and K Zhao. High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4):823–837, May 2007.

[172] Wikipedia. Signal-to-noise ratio - wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Signal-to-noise_ratio. [Online; accessed 31-March-2012].

[173] C Gambacorti-Passerini. Part i: Milestones in personalised medicine–imatinib. *Lancet Oncol*, 9(6):600–600, Jun 2008.

[174] 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 Gu, R D Polakiewicz, J Rush, and M J Comb. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell*, 131(6):1190–1203, Dec 2007.

[175] D T Yeung and T P Hughes. Therapeutic targeting of bcr-abl: Prognostic markers of response and resistance mechanism in chronic myeloid leukaemia. *Crit Rev Oncog*, 17(1):17–30, 2012.

[176] Y J Bang. The potential for crizotinib in non-small cell lung cancer: a perspective review. *Ther Adv Med Oncol*, 3(6):279–291, Nov 2011.

[177] G K Geiss, R E Bumgarner, B Birditt, T Dahl, N Dowidar, D L Dunaway, H P Fell, S Ferree, R D George, T Grogan, J J James, M Maysuria, J D Mitton, P Oliveri, J L Osborn, T Peng, A L Ratcliffe, P J Webster, E H Davidson, L Hood, and K Dimitrov. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*, 26(3):317–325, Mar 2008.

[178] nanoString Technologies. ncounter - gx human kinase kit. http://nanostring.com/uploads/nCounter_GX_Human_Kinase_Kit_PDS.pdf. [Online; accessed 5-April-2012].

[179] J. Sambrook and D.W. Russell. *Molecular cloning: a laboratory manual*, volume 2. CSHL press, 2001.

[180] P Fortina and S Surrey. Digital mrna profiling. *Nat Biotechnol*, 26(3):293–294, Mar 2008.

[181] nanoString Technologies. ncounter - gx human kinase system. http://nanostring.com/. [Online; accessed 5-April-2012].

[182] A A Alizadeh, M B Eisen, R E Davis, C Ma, I S Lossos, A Rosenwald, J C Boldrick, H Sabet, T Tran, X Yu, J I Powell, L Yang, G E Marti, T Moore, J Hudson, L Lu, D B Lewis, R Tibshirani, G Sherlock, W C Chan, T C Greiner, D D Weisenburger, J O Armitage, R Warnke, R Levy, W Wilson, M R Grever, J C Byrd, D Botstein, P O Brown, and L M Staudt. Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769):503–511, Feb 2000.

[183] S Ramaswamy, K N Ross, E S Lander, and T R Golub. A molecular signature of metastasis in primary solid tumors. *Nat Genet*, 33(1):49–54, Jan 2003.

[184] R S Herbst. Review of epidermal growth factor receptor biology. *Int J Radiat Oncol Biol Phys*, 59(2 Suppl):21–26, 2004.

[185] S Couraud, G Zalcman, B Milleron, F Morin, and P J Souquet. Lung cancer in never smokers - a review. *Eur J Cancer*, Mar 2012.

[186] D M Jackman, V A Miller, L A Cioffredi, B Y Yeap, P A Jänne, G J Riely, M G Ruiz, G Giaccone, L V Sequist, and B E Johnson. Impact of epidermal growth factor receptor and kras mutations on clinical outcomes in previously untreated non-small cell lung cancer patients: results of an online tumor registry of clinical trials. *Clin Cancer Res*, 15(16):5267–5273, Aug 2009.

[187] K Suda, K Tomizawa, and T Mitsudomi. Biological and clinical significance of kras mutations in lung cancer: an oncogenic driver that contrasts with egfr mutation. *Cancer Metastasis Rev*, 29(1):49–60, Mar 2010.

[188] M Soda, Y L Choi, M Enomoto, S Takada, Y Yamashita, S Ishikawa, S Fujiwara, H Watanabe, K Kurashina, H Hatanaka, M Bando, S Ohno, Y Ishikawa, H Aburatani, T Niki, Y Sohara, Y Sugiyama, and H Mano. Identification of the transforming eml4-alk fusion gene in non-small-cell lung cancer. *Nature*, 448(7153):561–566, Aug 2007.

[189] P M Ellis, N Blais, D Soulieres, D N Ionescu, M Kashyap, G Liu, B Melosky, T Reiman, P Romeo, F A Shepherd, M S Tsao, and N B Leighl. A systematic review and canadian consensus recommendations on the use of biomarkers in the treatment of non-small cell lung cancer. *J Thorac Oncol*, 6(8):1379–1391, Aug 2011.

[190] A Ciccodicola, R Dono, S Obici, A Simeone, M Zollo, and M G Persico. Molecular characterization of a gene of the 'egf family' expressed in undifferentiated human ntera2 teratocarcinoma cells. *EMBO J*, 8(7):1987–1991, Jul 1989.

[191] R Brandt, N Normanno, W J Gullick, J H Lin, R Harkins, D Schneider, B W Jones, F Ciardiello, M G Persico, and F Armenante. Identification and biological characterization of an epidermal growth factor-related protein: cripto-1. *J Biol Chem*, 269(25):17320–17328, Jun 1994.

[192] R C Doebele, A B Pilling, D L Aisner, T G Kutateladze, A T Le, A J Weickhardt, K L Kondo, D J Linderman, L E Heasley, W A Franklin, M Varella-Garcia, and D R Camidge. Mechanisms of resistance to crizotinib in patients with alk gene rearranged non-small cell lung cancer. *Clin Cancer Res*, 18(5):1472–1482, Mar 2012.

[193] B Valsasina, I Beria, C Alli, R Alzani, N Avanzi, D Ballinari, P Cappella, M Caruso, A Casolaro, A Ciavolella, U Cucchi, A De Ponti, E Felder, F Fiorentini, A Galvani, L M Gianellini, M L Giorgini, A Isacchi, J Lansen, E Pesenti, S Rizzi, M Rocchetti, F Sola, and J Moll. Nms-p937, an orally available, specific small-molecule polo-like kinase 1 inhibitor with antitumor activity in solid and hematologic malignancies. *Mol Cancer Ther*, Mar 2012.

[194] E Raymond, S Faivre, and J P Armand. Epidermal growth factor receptor tyrosine kinase as a target for anticancer therapy. *Drugs*, 60 Suppl 1:15–23, 2000.

[195] P Gresner, J Gromadzinska, and W Wasowicz. Reference genes for gene expression studies on non-small cell lung cancer. *Acta Biochim Pol*, 56(2):307–316, 2009.

[196] P A Nguewa, J Agorreta, D Blanco, M D Lozano, J Gomez-Roman, B A Sanchez, I Valles, M J Pajares, R Pio, M J Rodriguez, L M Montuenga, and A Calvo. Identification of importin 8 (ipo8) as the most accurate reference gene for the clinicopathological analysis of lung specimens. *BMC Mol Biol*, 9:103–103, 2008.

[197] J Vandesompele, K De Preter, F Pattyn, B Poppe, N Van Roy, A De Paepe, and F Speleman. Accurate normalization of real-time quantitative rt-pcr data by geometric averaging of multiple internal control genes. *Genome Biol*, 3(7), Jun 2002.