GENOME-WIDE LOCALISATION ANALYSIS FOR IRF4 TARGET GENE IDENTIFICATION IN MELANOMA CELL LINES

by

Ulduz Sobhi Afshar

B.S., Cellular and Molecular Biology, Tehran University, 2011

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Molecular Biology and Genetics

Boğaziçi University

2015

ACKNOWLEDGEMENTS

My utmost thanks and sincere gratitude and respect to my supervisor, Asst. Prof. Tolga Emre, whose continuous support and patience inspired me all the way and made this project possible. Your confidence in me and constitutive criticism to this research were instrumental and encouraging. Your optimism and persistence have pushed me in ways I could never have done for myself. I would never find the right word to thank you for trusting me with this awesome project and helped me to set it up from almost a scratch. I extend my thanks and gratitude to Assoc. Prof. Nesrin Özören and Asst. Prof. Mustafa Baydoğan, who stood by me through this long and exciting process. Your honesty, understanding, patience and persistence have been invaluable. In your own ways, you have each added to this project as well as helped to shape me as a researcher.

Much gratitude goes to Hengameh Mohammad Nezami, Mehrnaz Salehian and Ghazal Hasani and Sara Bagheri for their unwavering support and friendship, Erdem Yılmaz, Mustafa Can Ayhan, Cansu Yerinde, Ahmet Buğra Tufan, Ekin Ece Erkan and Stefan Köstler for their constant encouragement and support, and Nalan Yıldız and Gizem Gül for their lasting friendship and their constructive criticism they gave me at every opportunity. I would also like to thank İbrahim Taştekin and Erdem Yılmaz, particularly in desperate times unknowingly were a driving force in my work. All graduate students of MBG department, you all inspired me in your own way, your words and humor has kept me sane in desperate times. Last but not least, special thanks to all my friends who helped me with sending articles, protocols, scientific advices from all around the world. It was an honor and a privilege to have all of you in my life. Along the way, I have received support from too many people to count. Though you may not see your names here, in black and white, know that your various contributions have not gone unnoticed or unappreciated.

ABSTRACT

GENOME-WIDE LOCALISATION ANALYSIS FOR IRF4 TARGET GENE IDENTIFICATION IN MELANOMA CELL LINES

Interferon Regulatory Factor 4 (IRF4) is a key transcription factor in development and function of immune cells. Also it has been shown that elevated IRF4 expression is a key factor for survival in some myeloid and lymphoid cancers. Recently, studies have shown IRF4 expression also in non-immune cells and malignancies such as melanocytes and melanoma. Studies from our lab and elsewhere demonstrated elevated expression levels of IRF4 is critical for melanocytes and melanoma cell lines. There is lack of knowledge about IRF4 regulated genes in melanoma, so we aimed at identification of its localization genome-wide using high-throughput sequencing of immunoprecipitated chromatin (ChIP-seq). First, we established and optimized ChIP-qPCR (Chromatin immunoprecipitation coupled with quantitative polymerase chain reaction) experiments in our lab and confirmed several IRF4 binding regions in different loci such as Tyrosinase, a key developmental gene in melanocytes. Then, we performed ChIP-seq to identify IRF4 target genes genome-wide in a melanoma cell line. We set up a bioinformatics pipeline and analysed the ChIP-seq data to find putative IRF4 binding regions and their associated genes. Afterwards, we characterized DNA motifs,associated genes, pathways and biological processes. Our results integrated with previous RNA-seq data suggests that IRF4 is regulating genes related to cell cycle and proliferation, and in cellular transport system.

ÖZET

IRF4 HEDEF GENLERİNİN MELANOM HÜCRE HATLARINDA TANIMLANMASI AMACIYLA GENOM DÜZEYİNDE LOKALİZASYON ANALİZİ

İnterferon düzenleyici faktör 4 (IRF4), bağışıklık sistemi hücrelerinin fonksiyonlarında ve gelişimde önemli rol oynar. Ayrıca, IRF4 anlatımındaki artışın, bazı miyeloid ve lenfoid kanserlerinin hayatta kalmasında etkili bir faktör olduğu gösterilmiştir. IRF4 anlatımının melanosit ve melanom gibi immünolmayan hücrelerde ve tümörlerde de mevcut olduğu gösterilmiştir. Bizim laboratuvarımızda ve başka yerde yapılan araştırmalar, yüksek seviyedeki IRF4 anlatımının melanosit ve melanom hücre hatlarının yaşamsal faaliyetleri için kritik olduğunu göstermiştir. Melanomda IRF4 tarafından düzenlenen genler konusunda bildiklerimiz oldukça kısıtlı olduğu için Kromatin İmmünçoktürme-yüksek kapasiteli DNA dizilemesi (ChIP-seq) kullanarak IRF4'ün genom düzeyindeki lokalizasyonunu tanımlamayı hedefledik. Laboratuvarımızda ilk olarak ChIP-kantitatif PCR (ChIP-qPCR) methodunu kurup optimize ederek farklı bölgelerde birkaç IRF4 bağlanma bölgesi tespit ettik. Melanosit gelişimi için önemli bir gen olan tirozinaz bu bölgelerde bir örnek tir. Sonrasında, genom düzeyindeki hedef bölgeleri belirleyebilmek için bir melanom hücre hattında ChIP-seq uyguladık. Biyoinformatik yöntemler ile ChIP-seq verisini analiz ederek genome üzerinde IRF4 bağlanması gösteren bölgeleri, DNA bağlanma motifelerin ve ilişkili genleri bulduk. Sonraki adımda ilişkili genlerin, yolakların ve biyolojik süreçlerin, tanımlanmasını yaptık. Önceki RNA-seq verileri ile birleştirildiğinde bizim sonuçlarımız, IRF4'un hücre döngüsü ve proliferasyon ile ilişkili genleri, ve hücre içi taşıma sistemini düzenlediğini önermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$^{\circ}C$	Celsius
bp	Base Pairs
cm	Centimeter
gr	Gram
Kb	Kilobase
kDa	Kilodalton
L	Liter
M	Molar
mA	Milliamper
mg	Milligram
min	Minute
ml	Milliliter
mm	Millimeter
mM	Millimolar
ng	Nanogram
s	Second
V	Volt
μg	Microgram
μl	Microliter
α	Alpha
β	Beta

LIST OF ACRONYMS/ABBREVIATIONS

$^{\circ}C$	Celsius
cm	Centimeter
CO_2	Carbon dioxide
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetate
FBS	Fetal Bovine Serum
gDNA	Genomic DNA
gr	Gram
IMDM	Iscove's Modified Dulbecco's Medium
kDa	Kilodalton
mg	Milligram
ml	Milliliter
mm	Millimeter
min	Minute
ng	Nanogram
nm	Nanometer
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
rpm	Revolution per minute
RT	Room Temperature
q-PCR	Quantitative PCR
NaCl	Sodium chloride
SDS	Sodium dodecyl sulfate
u	Unit
UV	Ultraviolet

V	Volt

1. INTRODUCTION

1.1. Overview of Cancer

Cancer is caused by irregular and unrestrained growth and proliferation of cells which is consequence of accumulation of genetical and epigenetic alterations in the genome. The cells evade cell cycle controls, cell death and senescence, and initiate invasion and metastasis. In Figure 1.1, Hanahan and Weinberg's hallmarks of the cancer are shown [1] .

Figure 1.1. Hallmarks of Cancer (Adapted from [1]).

The most vital and key ability of a cancer cell is to sustain constitutive proliferation and skip its negative regulators. Cancer cells deregulate growth and proliferation promoting signals. One of the ways to achieve unregulated proliferation is to produce

growth factor ligands themselves. Also another way to promote mitogenic signaling is through increasing receptor protein levels and altering their structure, so they will be constitutively active [1]. One of the most famous deregulated receptors is Receptor Tyrosine Kinase (RTK) which becomes active and fire signals whether a ligand is present or not. This constitutive activity is due to a somatic mutation which alters RTK structure in a way that it is activated independent of its ligand which are growth signal molecules. Then Ras signaling pathway gets activated which can activate Myc, one of the key oncogenes in various cancers, therefore Ras pathway constitutes one of the main players of proliferation circuit [7].

There are different intrinsic cellular mechanisms in order to control proliferation rates. One of these ways is the use of negative feedback mechanisms that diminish proliferation promoting signals. Cancer cells need to deregulate and inactivate these pathways whether through suppressing the related genes or altering structure of the functional protein. In cancer cells these negative signals are diminished, thus enhancing proliferation signals. For instance, PTEN phosphatase, whose activity leads to negative regulation of proliferation, degrades phosphatidylinositol (3, 4, 5) triphosphate (PIP3) in PI3K signaling pathway. In cancer cells, there are loss-of-function mutations or DNA hypermethylation at the promoter of PTEN, either of which leads to its inactivation. Consequently, PTEN inactivation will increase PI3K signaling and promote tumorigenesis [8]. Another example of intrinsic cellular defensive mechanism is senescence. In healthy cells, if excessive proliferation signals are induced, cells go through senescence, which is a state of the cells in which they cease dividing and their morphology changes. During senescence cells show flattened shapes and their senescence-associated β-galactosidase staining assays are positive. In order to avoid senescence and cell death, cancer cells either have to compromise between the maximum amount of mitogenic signals to continue proliferation while at the same time avoid senescence, or they should adapt to the conditions of increased levels of mitogenic stimulation through elevated amount of oncoproteins like Raf, Myc and Ras and also inactivate their senescence, and even apoptosis signals. As a result the cancer cells would survive and continue proliferation.

In healthy cells, when there is DNA damage or defect in cell cycle, cells with the help of numerous proteins such as some tumor suppressor genes activate the apoptosis, temporary cell cycle arrest, or senescence pathway. Other common key tumor suppressor genes that are deactivated in cancer cells are TP53 and RB (retinoblastoma-associated). Both of these proteins, when the conditions for progression of cell cycle into growth and proliferation are not given, take action and call a halt in the cell cycle. Unfavorable conditions can either be due to environmental stress such as hypoxia, ultraviolet radiation or to genome damages. Cancer cells need to circumvent the effect of these proteins on proliferation through inactivating them via deletion or loss-of-function mutations [9, 10].

The next step for cancer cells is invasion and metastasis into other tissues. For invasion and metastasis cancer cells need to interact with other cells in their tumor microenvironment in order to acquire the capability of invasive growth. Usually this event happens through epithelial-mesenchymal transition (EMT). One of the key steps of EMT is diminishing cell-cell adhesion. Cancer cells in crosstalk with adjacent cells, and with the help of proteases in the stroma, degrade cell-cell adhesion related proteins like E-cadherin in order to facilitate invasion and migration. In cancer cells, E-cadherin is downregulated and sometimes it is inactivated through mutation [11]. To promote EMT, carcinoma cells activate various EMT-related transcription factors such as Snail, Slug, Twist, and Zeb1/2. Dependent on tumor type, different sets of these proteins are expressed in order to program and orchestrate invasion and metastasis [12]. By alteration and deregulation of different proteins involved in cell-cell adhesion and with the help of their tumor-microenvironment and adjacent cells, carcinoma cells promote their migratory and invasive capabilities in the invasion-metastasis cascade and start migration through the circulatory system to distant sites and invade other tissues and organs.

1.2. Melanoma

1.2.1. Overview of Melanoma

Melanoma is one of most aggressive and lethal kind of skin cancer with increasing incidents worldwide. Melanoma is causing about 75% of deaths related to skin cancer and with an incidence of 15–25 per 100,000. Despite efforts for early detection and prevention, approximately 20% of patients with melanoma die from their disease [13]. The underlying reason on high lethality of melanoma relative to other skin cancers is related to its high metastatic potential and resistance to therapy.

Cutaneous subtype of melanoma is caused by the transformation of skin melanocytes (pigment-producing cells) accumulating genetic and epigenetic alterations. The transformation of melanocytes start with formation of nevi which are benign neoplasms. Then the cells with accumulating genetical alterations acquire horizontal growth ability, or as it is called radial growth phase (RGP) of melanoma. This state is the primary malignant phase. Next step is the growth in vertical manner, the vertical growth phase (VGP) 1.2. In this stage melanoma cells will attain the ability to migrate and initiate invasion and metastasis [2].

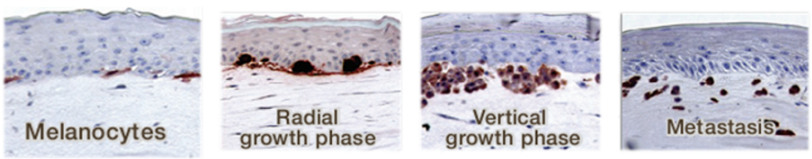


Figure 1.2. From melanocytes to melanoma. It starts with nevus and then transforms into radial growth phase and then goes to vertical growth phase and proceed with invasion and metastasis (Adapted from [2]).

1.2.2. Common Deregulated Pathways of Melanoma

In tumorigenesis and progression of cancer, various genetic and epigenetic alterations are involved in proliferation and metastasis (Section 1.1, [1]). Also in melanoma

genetic abnormalities in different pathways such as MAPK and PI3K pathways are accountable for the transformation of melanocytes to melanoma cells. Deregulation and over-activation in such pathways leads to abnormal proliferation and growth phase and eventually it can cause metastasis.

MAPK (Mitogen-activated protein kinase) pathway is responsible for regulation of cell proliferation, invasion, migration and survival in melanoma cells [14]. Alterations in membrane receptors or mutations in RAS or BRAF will lead to MAPK signaling pathway to be constitutively activated. Mutation in BRAF is seen in about 60% of melanoma patients, with a valine to glutamic acid substitution (V600E), triggering ligand-independent kinase activation. In 20% of melanomas, a point mutation of NRAS occurs which causes substitution of Q61R (glutamine to arginine) or Q61K (glutamine to lysine) [15]. Mutations in NRAS make it constitutively active. BRAFV600E is the hyperactive form of BRAF with increased kinase activity. Downstream targets of these proteins are MEK1/2 and ERK1/2 proteins which are known activating signals for transcription factors related with proliferation, survival and tumor growth [16, 17].

NRAS and BRAF mutations are not sufficient for the progression of melanoma so their alteration will combine and accumulate with other genetic and epigenetic changes in the genome such as inactivation of PTEN,tumor-suppressor gene, in order to activate the invasion-metastasis cascade [17, 18].

Despite the lower mutation frequency in melanoma patients compared to BRAF and MAPK pathways, PI3K signaling pathway is one of the major activated pathways in different melanoma tumor samples and has various downstream effectors such as AKT, mTOR, NF-κB, p53, and others, all possibly leading to a more aggressive cancer phenotype. Although PTEN is mutated in 4% of all melanoma samples, studies show that in about 44% of BRAF-mutated melanoma samples PTEN mutation, gene deletion or promoter methylation are present. Loss of PTEN will lead to activation of PI3K signaling pathway. The downstream targets of PTEN and PI3K signaling pathway are transcription factors such as microphthalmia-associated transcription factor (MITF) which is master regulator and one of the key genes in melanocyte development. As it

is shown in Figure 1.3 . PI3K pathway downstream effectors are involved in cell growth and survival [17].

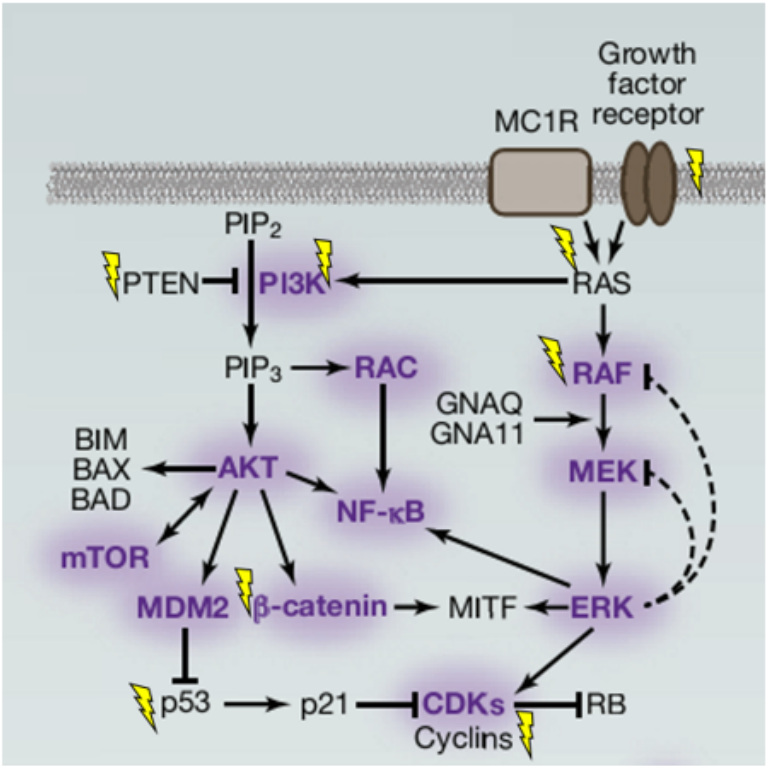


Figure 1.3. Common deregulated pathways in melanoma (Adapted from [2]). Yellow thunderbolt shapes show the common altered proteins and violet highlight designate the proteins with potential of being targeted by drugs.

Another important gene in melanoma which gets inactivated is CDKN2A. Inactivation of CDKN2A important for cells to avoid senescence. CDKN2A is a tumor-suppressor gene which encodes for two different cyclin-dependent kinase inhibitor proteins through alternative splicing: $p16^{INK4a}$ which negatively regulates CDK4 and therefore activation of Rb protein, and $p14^{ARF}$ which is responsible for P53 activation via inhibition of MDM2 [16, 19]. Loss of $p14^{ARF}$ gene causes a decrease in P53 levels in the cell and thus, leads to cell survival. Previous studies have demonstrated combining $NRASQ^{61K}$ mutation and loss of $p16^{INK4a}$ is adequate to stimulate melanoma formation in mice [20]. The master regulator of melanocyte development, Microphthalmia-associated transcription factor (MITF), plays critical role in melano-

genesis and melanin synthesis. Downstream targets of MITF are genes such as Tyrosinase (TYR), Tyrosinase-related protein 1 (TRYP1) and Dopachrome tautomerase or tyrosine-related protein 2 (DCT) [21]. Recent studies have shown that MITF also plays part in proliferation and survival of melanoma cells [22,23]. Another recent study demonstrated that during melanoma progression, depending on the phenotype of the melanoma cell, MITF levels are changing [3]. Elevated levels of MITF lead to cell proliferation at least by activating a gene called DIAPH1 [24]. In addition, high MITF levels suppress BRN2 expression through transcriptional activation of miR-211. BRN2 is a transcription factor, adept of amplifying the invasive potential of melanoma cells in vitro [25]. On the other hand, low levels of MITF is implicated in the invasive and anti-proliferation phenotype via the activity of ROCK (Rho-associated protein kinase) [24]. Figure 1.4 demonstrates the alternation of two phenotypes based on MITF expression.

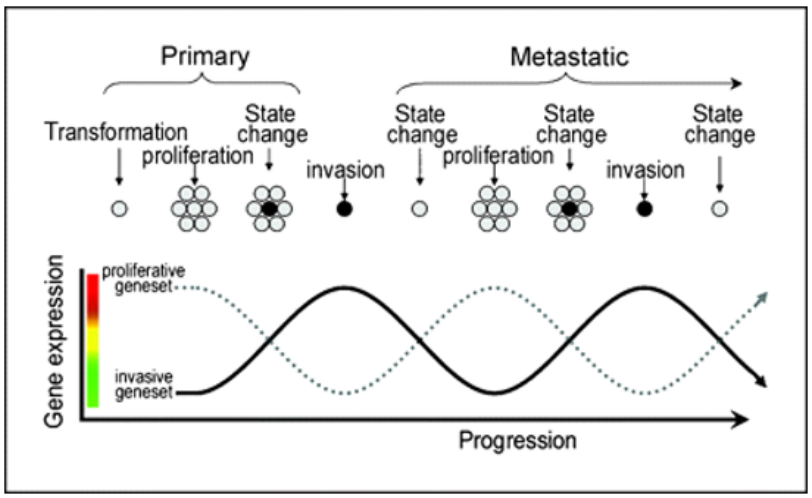


Figure 1.4. Phenotype switching model of melanoma (Adapted from [3]).

1.3. Interferon Regulatory Factor 4 (IRF4)

1.3.1. Overview of IRF4 and its Role in Development and Disease

Interferon regulatory factor 4 (IRF4) protein is a transcription factor from IRF (Interferon regulatory factors) family. Throughout literature IRF4 has different names

such as MUM1 (Multiple Myeloma Oncogene 1), LSIRF (Lymphocyte specific interferon regulatory factor), PIP (PU.1 interaction partner) or ICSAT (Interferon consensus sequence binding protein for activated T-cells). IRF4 can function either as an activator or repressor of gene transcription in conjunction with numerous cofactors [26].

One of the first studies of IRF4 was in B cells as transcriptional regulator in immunoglobulin production and B-cell differentiation. In this study, IRF4 knockout mice showed immune deficient characteristics such as abnormal lymph node and spleen size, accumulation of immature B cells, and reduced serum concentrations of all IgG subclasses [27].

In cells of hematopoietic origin most IRF family members are induced by interferons. IRF4, one of the exceptions, is induced by various mitogenic stimuli. These stimuli include: antigen receptor engagement, lipopolysaccharides and CD40 signaling and cytokine signaling. The common point between all these stimuli is that they activate the NF-κB pathway which leads to activation of the IRF4 promoter via NF-κB heterodimers. Additionally, IL-4 cytokine involving STAT6 transcription factor can activate IRF4 [27, 28].

Depending on the lineage and stage of development of hematopoietic cells, variable levels of IRF4 expression were observed. Mitf transcription factor represses IRF4 expression in naïve B cells leading to prevention from spontaneous plasma cell differentiation. Consequently, when Mitf function is defective, IRF4 gets activated and overexpressed which leads to spontaneous plasma cell differentiation [29, 30].

IRF4 is also a key factor in T-helper cell and myeloid-cell differentiation. In development of Th2 and Th17 type T-helper cell family, IRF4 functions as regulator of cytokine expression and apoptosis [31, 32].

IRF4, because of its auto-inhibitory C-terminal domain, has weak DNA-binding ability by itself. However, in B cells and T cells, IRF4 cooperates with PU.1 or SPIB of ETS family transcription factors, leading to increased IRF4-DNA binding [33, 34].

When IRF4 partners with PU.1 or SPIB, they bind to ETS/ISRE-consensus element (EICE), which has consensus sequence of $5' - GGAANNGAAA - 3'$. This sequence includes two different motifs, the ETS binding motif ($5' - GGAA - 3'$) and IRF binding motif ($5' - AANNGAAA - 3'$) [26,35]. Cooperative binding of PU.1 and IRF4 increases IRF4-DNA binding tendency by 5 folds [36]. When IRF4 interacts with E47, an E-box binding transcription factor, their DNA binding tendency can increase up to 50-100 fold [37]. In T cells, it has been shown that cooperation of IRF4 and NFATC2 regulates IL-4 promoter [38]. A more recent study has demonstrated that in mouse $CD4^+$ T cells, IRF4 and Activator Protein 1 (AP-1) co-operate and binds to AP-IRF composite elements (AICEs) motifs ($5' - TGAnTCA/GAAA - 3'$). BATF-JUN protein complex family interacts with IRF4 in binding to AICE in preactivated $CD4^+$ T cells and TH17 differentiated cells. One of the target genes of AP-1 and IRF4 cooperation is IL-10 [39].

Considering IRF4's key role in survival and proliferation of hematopoietic cells during development and differentiation, it might have been expected that IRF4 also would have a key role in some of hematopoietic malignancies. One of the first studies on the role of IRF4 in malignancies was in 1997 in multiple myeloma (MM), a plasma-cell related cancer. In this cancer, because of the translocation (6; 14) (p25; q32) immunoglobulin heavy chain locus is in the vicinity of IRF4 gene. Furthermore, they observed that IRF4 is overexpressed and IRF4 has oncogenic activity in vitro which may contribute to tumorigenesis [40].

Among various mature B-cell malignancies, IRF4 demonstrates elevated expression levels in activated B-cell-like diffuse large B-cell lymphomas (ABC-DLBCL). ABC-DLBCL cells are results of transformation of plasmablasts, which correspond to the post germinal center differentiation step of activated B-cells. One of key facts of ABC-DLBCL cancer is its constitutive NF-κB pathway activation. Similar to differentiation of B-cells, in ABC-DLBCL malignancy high IRF4 expression is regulated by the NF-κB pathway. On the other hand, in GCB-DLBCLs (Germinal center B-cell like DLBCL), because of low activity of NF-κB pathway, there is no detectable IRF4 expression. In ABC-DLBCLs, IRF4 is not the only highly expressed transcription factor; in about 25% of ABC-DLBCL cases, SPIB, a partner of IRF4, levels are deregulated and highly

amplified, and works in conjunction with IRF4 [41–43].

However, the most studied role of IRF4 in malignancies is multiple myeloma (MM). IRF4 knockdown in MM cell lines demonstrated rapid non-apoptotic cell death in all MM cell lines tested [44]. There are several critical metabolic pathways controlled by IRF4 directly or by its downstream effectors such as: lipid and cholesterol biosynthesis, glucose metabolism, and cell cycle progression. One of IRF4 key targets in MM is MYC. IRF4 binds to the MYC promoter in MM cells and activates MYC. Moreover, MYC, through a conserved intronic region, regulates IRF4 expression in a positive autoregulatory feedback loop. MM cells have non-oncogenic addiction toward IRF4. Non-oncogenic addiction means that MM cells, due to their genetical abnormalities, are highly dependent on a normal cellular protein which in this case is IRF4 [45].

Controversial to the role of IRF4 in MM and ABC-DLBCLs, in Chronic myeloid leukemia (CML, transformation of granulocytes) and B cell acute lymphoblastic leukemia (B-ALL, lymphoblast origin), acute myeloid leukemia (AML, originated from myeloblasts) and chronic myelomonocytic leukemia (CMML, originated from hematopoietic stem cells) IRF4 plays tumor-suppressor roles [46–48]. This situation is a result of difference between roles of IRF4 in immature immune cells verses mature immune cells. It is likely this originates from the fact that expression of IRF4 during early stages of B cell development results in cell-cycle arrest and growth inhibition [49].

IRF4 also has roles in myeloid lineage in particular differentiation of dendritic cells (DC) and macrophages. IRF4 mRNA levels are elevated during differentiation of monocytes into macrophages or dendritic cells [50]. M2 macrophages (alternatively activated macrophages) can be activated during TH2-type responses which is characterized by increases in IL-4 and other TH2 cytokines. Irf4, alongside NF-κB, JAK-STAT pathway and others, play crucial role in the induction of M2 macrophage responses [51]. IRF4 plays a key role in development of various dendritic cell populations. IRF4 deletion in DC results in decreased development and survival of CD103+CD11b+ dendritic cells which leads to defective Th17 responses [52, 53]. IRF4 also has role in modulating Toll-like-receptor signaling (TLR) in dendritic cells. In dendritic cells, for Th2 response

promotion, IRF4-activated IL-10 and IL-33 play a pivotal role [54, 55].

For a long time, it was considered that IRF4 expression is restricted to hematopoietic cells such as B lymphocytes, T lymphocytes, dendritic cells and macrophages. Recently, it was discovered that IRF4 is expressed and has a functional role also in non-hematopoietic origin cells. IRF4 expression has been observed in adipocytes, cardiac tissue, central nervous system and melanocytes. In adipocytes, IRF4 stimulates lipolysis, a process in which fat cells break down triglycerides into free fatty acids and glycerol to produce energy during fasting conditions [56]. In a recent study, it has been shown that expression of IRF4 promotes cardiac hypertrophy, which is a condition where the heart muscle is thickened because of pressure [57]. In CNS, Guo et al. observed that high levels of IRF4 plays a protective role in neuronal survival through preventing apoptosis and cell-death during ischemia and reperfusion [58].

1.3.2. IRF4 in Melanocytes and Melanoma

The first lines of evidence about expression of IRF4 expression in melanocytes was shown in a normal foreskin melanocyte expressed sequence tag library (EST library) [59]. Also high IRF4 mRNA expression levels is observed in melanomas in different databases such as Oncomine [60] or Cancer Cell Line Encyclopedia [61]. In Figure 1.5, it is clear that the only cells in epidermis expressing IRF4 mRNA are Melanocytes [62].

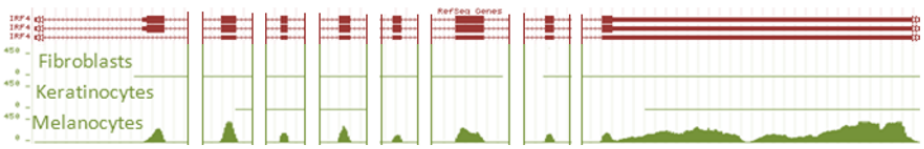


Figure 1.5. IRF4 expression three different cell types, Fibroblasts, Keratinocytes and melanocytes. Origin of this RNA-Seq data is from NIH Roadmap Epigenomics Project. Intronic regions were shortened for a better view. The three rows at the top shows three different IRF4 transcripts. Thick red bars represent exons. Visualization is from UCSC Genome Browser.

In genome-wide association studies (GWAS) in melanocytes, it has been demonstrated that a single nucleotide polymorphism (SNP) in IRF4 fourth intronic region (rs12203592) is associated with pigmentation phenotypes such as blue eyes, brown hair, freckles and sun sensitivity. This SNP lies within an enhancer for IRF4 transcription in melanocytes. Individuals with rs12203592 variant allele have high IRF4 expression [63, 64].

In a recent study it has been demonstrated that MITF, cooperating with IRF4, activates TYR (tyrosinase) expression in melanocytes which has a key role in melanogenesis and melanocyte differentiation. Regarding rs12203592 SNP, the IRF4 levels in homozygous C/C individuals are higher at both mRNA, as well as protein levels, compared to homozygous T/T individuals. The rs12203592-C variant site harbors the binding site for TFAP2A. As it is shown in Figure 1.6, in this site TFAP2A cooperates with MITF to induce IRF4 expression [4].

The first evidence of IRF4 expression in melanoma was observed in G-361 cell line [59]. Natkunam et. al study showed high IRF4 expression at protein level by immunohistochemistry in melanoma patient samples [65]. In a later study, due to positive staining for IRF4 in 92% of primary and metastatic melanoma patient samples, it was suggested that IRF4 may be a potential diagnostic marker [66]. Moreover from GWAS studies, it has been discovered that the germline SNP rs12203592 (described above) in IRF4 predisposes individuals to melanoma and other skin cancers [63,67,68]. However, the mechanisms underlying IRF4 expression and its role in melanoma survival is yet to be determined. In this project, we set out to shed some light on identification of IRF4 target genes genome-wide in melanoma cell lines by high throughput sequencing of chromatin immunoprecipitated samples (ChIP-seq).

1.4. A Brief Overview of Protein-DNA Localization Studies

Proteins interact with DNA in various ways such as: hydrogen bonds, salt bridges, and hydrophobic interactions. All these different ways of interaction and binding may contribute to sequence specificity of nuclear protein-DNA interactions, which are crit-

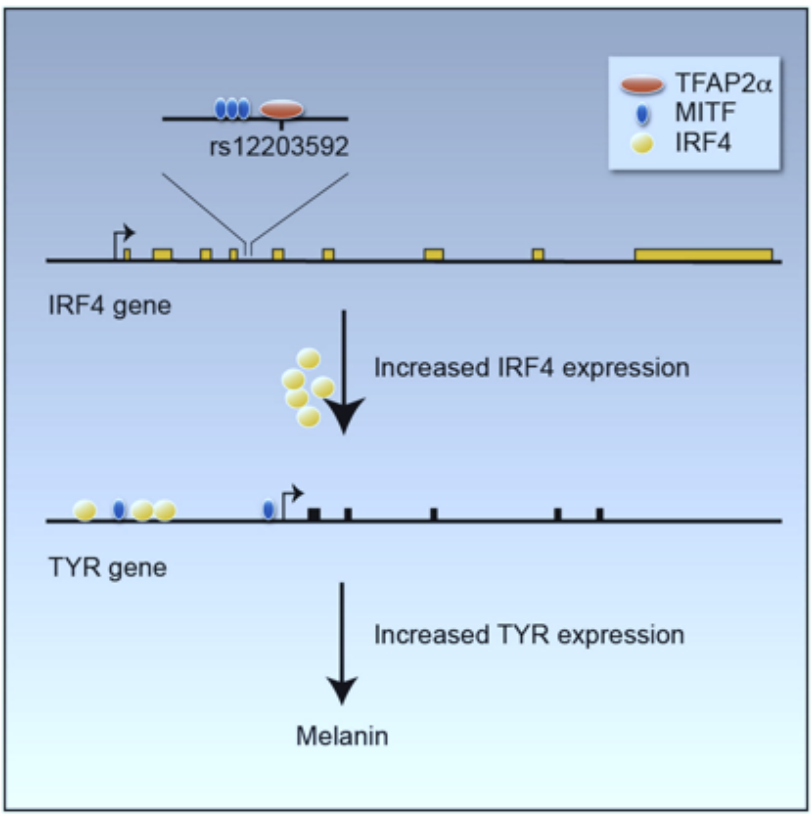


Figure 1.6. Cooperation of IRF4, MITF and TYR in melanocytes (Adapted from [4]).

ical for all cells. Many biological processes such as genome packaging, replication and DNA repair, and gene transcription are depended on these interactions. Therefore, study of protein-DNA interaction is essential in order to elucidate underlying mechanisms in biological process such as growth, development and differentiation. Additionally, since in cancer cells these biological process are vital, studying protein-DNA interaction will reveal in-depth details about deregulated genes and how their regulation have changed. In the last twenty years, different methods were developed to study protein-DNA interactions [69]. The most common methods are:

- Electrophoretic Mobility Shift Assay (EMSA)
- Chromatin Immunoprecipitation Assay (ChIP)
- DNA Pull-down Assay
- Immunofluorescence techniques and super-resolution microscopy
- Systematic evolution of ligands by exponential enrichment (SELEX)

EMSA is used in studying affinity or specificity of a candidate DNA-binding protein to known DNA oligo probes. This assay is based on the fact that free DNA molecules move faster through non-denaturing polyacrylamide (or agarose gel electrophoresis) than protein-DNA complexes. The DNA-binded protein complex will move slower and make a shift compared to only DNA sample, which is the reason for the other name of this assay, gel shift assay.

Another method, to study protein-DNA interaction is DNA pull-down assay. In this method a DNA probe is labeled with a high affinity tag, such as biotin. After treating this biotin-tagged DNA probe with cell lysate, via streptavidin magnetic beads the Protein-biotin-tagged-DNA is recovered. So protein-DNA complex is isolated and can be identified by mass spectrometry [69].

Chromatin immunoprecipitation (ChIP) is based on enrichment of DNA associated with the protein of interest, therefore the protein of interest target genes can be identified [70]. This method widely has been used for localization analysis of target genes in transcription factors, chromatin modifiers and histone marks [71]. Briefly, in

this method, cells are treated with cross-linking chemicals such as formaldehyde. Cell lysates containing covalently cross-linked protein-DNA complex are sonicated and then with the help of antibody-conjugated magnetic beads, desired protein-DNA complex are isolated an analyzed. This analysis can be performed via PCR, q-PCR locally and for just a few genes or genome-wide via high throughput methods like microarray and next-generation sequencing.

Systematic evolution of ligands by exponential enrichment (SELEX), which is also known as in vitro selection or in vitro evolution, is a technique in molecular biology but chemistry in nature. In this method, oligonucleotides are produced in DNA or RNA format in order to investigate if they precisely bind to a target ligand or ligands [72].

In vivo real-time visualization of protein-DNA interaction was a challenge for a long time. Recent developments in the field of microscopy, and super resolution microscopy techniques via fluorescent-tagged antibodies have contributed in clarification of more details about protein-DNA and protein-chromatin interactions and gene regulation in cells [73].

1.5. Next Generation Sequencing

The first developed nucleotide sequencing method is "Sanger sequencing". The mechanisms underlying this method is based on chain termination using dideoxynucleotides (ddNTPs) during in vitro replication [74]. ddNTPs, missing $3'$ -OH group, terminates the elongation of DNA. This reaction is catalyzed by DNA polymerase. A decade after invention of Sanger sequencing, first generation sequencing, modified version of Sanger was invented. In this method, with fluorescently-tagged ddNTPs information about the matching base is provided by running PCR products in capillaries (capillary electrophoresis) to sort the fragments according to size and order. Finally, depending on the fluorescent label, nucleotides are identified [75].

Sanger method is still widely used in sequencing of oligos and plasmids and genomic region for sequence confirmation. However for sequencing of large genomes like

eukaryotic genomes or transcriptomes, less time-consuming and cost-effective technology was a must. So next generation sequencing methods (NGS) was invented, and it can sequence millions of DNA fragments at the same time. In NGS method, detection of nucleotides is happening alongside their addition to the complementary strand whereas in Sanger method, nucleotide detection process happens with chain-termination, and separation every time a new nucleotide is added. At the moment, NGS is widely used for whole genome sequencing, de novo sequencing, targeted re-sequencing and exome sequencing for DNA, and mRNA or small RNA sequencing, and for sequencing of samples from ChIP [76].

Currently diverse technologies and techniques are applied in next generation sequencing systems such as sequencing by synthesis (Illumina), pyrosequencing (Roche/-454), and sequence by ligation (Life Tech/SOLiD). Each of these platform applies different strategies for library preparation, sequencing and imaging steps. In library preparation, bridge amplification or emulsion-PCR are used to amplify template DNA before sequencing. There are different sequencing approaches for each platform, fluorescent-labeled reversible terminator (Illumina), pyrosequencing (Roche/454), and cleavable probes which are sequenced by ligation (SOLiD). Because of different methodologies used at each step, each of these three different platforms has their own advantages and disadvantages. Roche/454 have the capability of generating relatively longer reads which can cause improvements in mapping of repetitive regions. Results from SOLiD are more accurate and error-free. However, Illumina platform still yields highest throughput per run [77, 78].

A promising novel technology is Nanopore sequencing (Oxford Nanopore Technologies). In this sequencing method, nanopores are formed from synthetic material or pore-forming proteins. These nanopores are the biosensors used to detect the nucleotides of a DNA strand as it passes through the pore. These nanopores are inside an insulating membrane separating the two chambers. When a DNA molecule passes through the pore, it causes disturbance in the ionic current inside the nanopore. Each nucleotide has its own characteristic change in the magnitude of current, so each nucleotide in a DNA strand can be detected while passing through the nanopore. Cur-

rently, researchers are exploring the potentials of this technology and developing its applications in sensing RNAs and proteins too [79, 80].

1.5.1. ChIP-seq

Chromatin Immunoprecipitation followed by sequencing is widely used to determine transcription factor binding sites (TFBS) genome-wide [81]. Before invention of NGS, and ChIP-seq, researchers used ChIP-on-chip method which is ChIP followed by microarray technology in order to find genome-wide TFBS [82]. Comparing ChIP-seq to ChIP-on-chip, the resolution of TFBS has been dramatically changed from a few hundred to tens of nucleotide. Recently, improvement of ChIP-seq, the ChIP-exo method, increased the TFBS resolution to single nucleotides [83].

There are several large-scale ChIP-seq projects are going on globally. Encyclopedia of DNA Elements (ENCODE) an international consortium which targets identification of all functional elements in the human genome, including regulatory elements in different cell lines. ENCODE currently has 457 ChIP-seq datasets on 119 TFs in a number of human cell lines [84]. Roadmap Epigenomics project, another international program which aims at developing standardized procedures for epigenomics research, develop new technologies for single cell epigenomic analysis and in vivo imaging of epigenetic activity and furthermore to create a public data resource of human epigenomic data to catalyze basic biology and disease-oriented research [62].

2. PURPOSE

Previous studies have established the foundation on role of Interferon Regulatory Factor 4 (IRF4) transcription factor in hematopoietic cell line development and lymphoid cancers. Genome-wide target genes of IRF4 protein have been identified in ABC-type diffuse large B-cell lymphomas (ABC-DLBCL) and multiple myeloma (MM) cell lines. Studies have shown IRF4 is one of key transcription regulators for the survival of multiple myeloma and ABC-DLBCL. In recent studies from our lab and elsewhere demonstrated elevated expression levels of IRF4 in melanocytes and melanoma cells, which encouraged us to investigate functional characterization of IRF4 in melanoma.

The purpose of this thesis is to identify genome-wide target genes of IRF4 in melanoma cell lines in order to elucidate target pathways and deregulated biological processes in human melanoma cell lines. For this purpose, we set up and optimized the chromatin immunoprecipitation experiments and validated the efficiency of ChIP with qPCR. After setting up and optimizing ChIP protocol, we performed ChIP-seq in order to do genome-wide localization analysis on IRF4 target genes. Furthermore, we analyzed the common genes of ChIP-seq and RNA-seq,in order to identify of IRF4-regulated genes and pathways which can help us to elucidate role of IRF4 in melanoma biology and offer a basis for designing and applying more accurate and effective treatment options in the long run.

3. MATERIALS

3.1. General Kits, Enyzmes and Reagents

Table 3.1. List of kits, enzymes and reagents.

DMEM	Gibco, LifeTechnologies, USA
FBS	Gibco, LifeTechnologies, USA
PCR Purification Kit	Macharey Nagel, Switzerland
	Qiagen, Germany
DNAstable Kit	Biomatrica, USA
Non Essential Amino Acids	Gibco, LifeTechnologies, USA
Penicillin / Streptomycin	Gibco, LifeTechnologies, USA
ChIP-DNA Isolation Kits	Zymo Research, USA
Protease Inhibitor Coctail Tablets	Roche, Switzerland
LightCycler® 480 High Resolution Melting Master	Roche, Switzerland
Luminaris Mastermix	Thermo Scientific, USA
Trypsin	Gibco, LifeTechnologies, USA
NEBNext® ChIP-Seq Library Prep Master Mix Set for Illumina®	New England Biolabs, USA
NEBNext® Ultra^{TM} DNA Library Prep Kit for Illumina®	New England Biolabs, USA
Proteinase K	Invitrogen, LifeTechnologies, USA
RNase A	Sigma Aldrich, USA
Agencourt AMPure XP	Beckman Coulter,USA
Agilent High Sensitivity DNA Kit	Agilent Technologies, USA
Dynabeads® Protein G for Immunoprecipitation	LifeTechnologies, USA
Dynabeads® Protein A for Immunoprecipitation	LifeTechnologies, USA

3.2. Biological Materials

3.2.1. Cell Lines

SKMEL-28, G-361 (human melanoma cell lines; kindly provided by Dr. María S. Soengas), A2058, SKMEL-5 (human melanoma cell lines; kindly provided by Yetiş Gültekin and Prof. David M. Sabatini), MCF-7 (human breast cancer cell line; kindly provided by Prof. Nesrin Özören), OCI-LY19 and HBL-1 (human ABC-DLBCL cancer cell lines; kindly provided by Prof. Georg Lenz), cell lines were used in the experiments.

3.2.2. Primers

Table 3.2. Primers used in this study.

Primer ID	Sequence	Application
EHBP1L1	Forward:5'-TCACGTTGGTACCAGGCAGCCCAA-3'	q-PCR
	Reverse: 5'-ACAGTGGCCGACTTTCAGTTCCCA-3'	
TFEB	Forward: 5'-AGGTCCCCAGAATCCAGCACCTCT-3'	q-PCR
	Reverse: 5'-ATGGACGTGGAAGGCCATGCTGA-3'	
SUB1	Forward:5'-CTTAGAGAACCGAAACCCAAACCTACA-3'	q-PCR
	Reverse: 5'-TGCAACCCTTCCTGCTTTAACAAGTTT-3'	
NEG1	Forward: 5'-ATCCCAAGTAGAACTGATAGCACCGTAA-3'	q-PCR
	Reverse: 5'-AGATGGTGGCCTCCCTTGTCTGCTGCTA-3'	
NEG2	Forward: 5'-AAACCAGGGCCGCACTAACAATGGTAA-3'	q-PCR
	Reverse: 5'-CATGAGGGACTGGCCTTTCTATAATTG-3'	
NEG3	Forward: 5'-AATATGTACATCAGGCAATCGGCTCTTC-3'	q-PCR
	Reverse: 5'-CAACTGGAATCAGATCCACTTCATGGAAA-3'	
pTYR	Forward: 5'- GTGGGATACGAGCCAATTCGAAAGA-3‘	q-PCR
	Reverse: 5'- CCCACCTCCAGCATCAAACACTTTT-3‘	
dTYR	Forward: 5'- TGTGTGGGTGAAGAGGAAGAGAAGT-3‘	q-PCR
	Reverse: 5'- TAAGCCTCCTTGTGGAGATCATGTG-3‘	

3.3. Chemicals

Table 3.3. Chemicals used in this study.

Ethidium Bromide	Merck, USA
EDTA	Merck, USA
Hydrochloric Acid	Merck, USA
Sodium Chloride	Merck, USA
Formaldehyde	Sigma Aldrich, USA
Tween 20	Merck, USA
Ethanol	Merck, USA
Triton-100	Merck, USA
Glycine	Sigma Aldrich, USA
DMSO	Merck, USA
Nonidet P40	Sigma Aldrich, USA
SDS	Applichem, Germany
	Sigma Aldrich, USA
Sodium Deoxycholate	Sigma Aldrich, USA
N,N'-Methylenbisacrylamide	Sigma Aldrich, USA
DEPC-treated water	Fisher Scientific, USA
Agarose	Sigma Aldrich, USA
Tris-Cl	Merck, USA
Tris-Base	Sigma Aldrich, USA
HEPES	Gibco Invitrogen, USA
NaCl	Merck, USA

3.4. Buffers and Solutions

Table 3.4. Buffers and solutions used in this study.

1.25 M Glycine	Distilled Water
ChIP Lysis Buffer	50mM HEPES
	150mM NaCl
	1% Triton-X 100
	0.1% Na-deoxycholate
	1mM EDTA
High Salt ChIP Lyis Buffer	50mM HEPES
	500mM NaCl
	1% Triton-X 100
	0.1% Na-deoxycholate
	1mM EDTA
ChIP Lysis Buffer with SDS	50mM HEPES
	150mM NaCl
	1% Triton-X 100
	0.1% Na-deoxycholate
	1mM EDTA
	0.25% SDS
Freezing Medium	1X DMEM
	20% FBS
	1X Pen/Strep
	100 μM MEM-NEAA
	10% DMSO

Table 3.5. Buffers and solutions used in this study (cont.).

Tris-EDTA	10 mM Tris, bring to pH 8.0 with HCl
	1 mM EDTA
1M HEPES	Distilled Water
10X PBS	150 mM M NaCl
	270 mM KCl
	80 mM NaH2PO4
	20 mM KH2PO4
1X PBS-T	50 mM Tris-Base (pH: 7.4)
	150 mM NaCl
	0.1% Tween 20

3.5. Antibodies

Table 3.6. Antibodies used in this study.

Name	**Species**	**Source**
Anti-IRF4 X (M-17)	Goat	Sc-6059 Santa Cruz
Anti-IRF4 X (N-18)	Mouse	Ab12039 Abcam
Normal Goat IgG	Goat	c-2028 Santa Cruz
Anti- IRF4	Rabbit	#4964 Cell signaling
Normal Rabbit IgG	Rabbit	#2729 Cell Signaling
NLRP7	Goat	Santa Cruz

3.6. Disposable Labware

Table 3.7. List of disposable lab wares used in this study.

Protein LoBind 2.0ml Tubes	Eppendorf, Germany
DNA LoBind 1.5ml Tubes	Eppendorf, Germany
Parafilm	Brand, Germany
Centrifuge Tubes, 15 ml	Vwr, USA
Centrifuge Tubes, 50 ml	Vwr, USA
Serological pipette, 5ml	Tpp, Switzerland
Serological pipette, 10ml	Tpp, Switzerland
Serological pipette, 25ml	Tpp, Switzerland
Pipette Tips, filtered	Biopointe, USA
Pipette Tips, bulk	Biopointe, USA
Microcentrifuge tubes	Axygen, USA
PCR Tubes, 0.2 ml	Axygen, USA
Medical Gloves	Vwr, USA
	Broche Medikal, Turkey
Syringe Filters	Sartorius, Germany
Cryovial	Tpp, Switzerland
Cell Culture Plates, 15 cm	Tpp, Switzerland
Cell Culture Flasks, 75cm2	Tpp, Switzerland
Glass Pasteur Pipette, 230 mm	Witeg, Germany
96 well plates for qRT-PCR	Thermo Scientific, USA

3.7. Equipments

Table 3.8. List of equipment used in this study.

Beaker, 100 ml	Duran, Germany
Beaker, 600 ml	Duran, Germany
Erlenmeyer, 250 ml	Duran, Germany
Erlenmeyer, 500 ml	Duran, Germany
Bottle, 1000 ml	Vwr, USA
Bottle, 500 ml	Vwr, USA
Bottle, 100 ml	Vwr, USA
Forceps	RSG Solingen, Germany
Magnetic Stirring Bar	Vwr, USA
Cryobox	Tenak, Denmark
Autoclave Indicator Tape	Llg, Germany
Measuring Cylinder 100 ml	Kartell, Italy
Measuring Cylinder 250 ml	Kartell, Italy
Measuring Cylinder 1000 ml	Kartell, Italy
pH Meter	Hanna, USA
Microtube Racks	Vwr, USA
Electronic Balance	Sartorius AY123, Germany
Microcentrifuge	Vwr Galaxy Ministar, USA
Horizontal Electrophoresis	Cleaver Scientific Multi Sub Mini, UK
Micropipettes	Cleaver Scientific, UK
	Gilson Pipetman Neo PCR Kit, USA
	Axygen, USA

Table 3.9. List of equipment used in this study (Cont.).

Heat Block	Cleaver Scientific EL-01,UK
Pipettor	Greiner Labopet 240, Germany
Vortex	Vwr, USA
Centrifuges	J2-21, Beckman Coulter, USA
	Allegra X-22, Beckman Coulter, USA
	5415R, Eppendorf, USA
Cell Culture Incubator	Binder C-150, Germany
Documentation System	GelDoc XR System, Bio-Doc, Italy
Freezers	-20°C Ugur UFR 370 SD, Turkey
	-80°C Thermo Scientific TS368, USA
	-150°C Sanyo MDF1156, Japan
Microscopes	Inverted Microscope Nikon Eclipse TS100, Japan
Thermal Cycler	Antarus MyCube ANT101, USA
	Biorad Thermal Cycler T100, USA
Power Supply	Vwr, USA
Real Time PCR Machine	Thermo Scientific Piko Real 96, USA
Spectrophotometer	NanoDrop 1000, USA
Stirrer - Heater	Dragonlab MS-H-S, China
Rotator	Grant Bio Multifunctional Rotator PTR-35, UK
Refrigerator	4°C Ugur USS 374 DTKY, Turkey
Refrigerated Vapor Trap	Thermo Scientific SPD111V, USA

Table 3.10. List of equipment used in this study (Cont.).

Oil-free Gel Pump	Thermo Scientific Savant VLP110, USA
Vacuum Pump Oil Filter	Thermo Scientific VPOF110, USA
SpeedVac	Thermo Scientific SPD111V, USA
Carbon dioxide Tank	Genç Karbon, Turkey
Ice Maker	Brema, Italy
	Scotsman Inc. AF20, Italy
Autoclave	Model ASB260T, Astell, UK
Dishwasher	Miele Mielabor G7783, Germany
Cold Room	Birikim Elektrik Soğutma, Turkey
Freezing Container	Nalgene, USA
Oven	Nüve KD200, Turkey
Rotator	HulaMixer Sample Mixer Life technologies, USA
Sonicator	Bandelin SonoPuls HD-2070, Germany
Magnetic Stand	PureProteome Magnetic Stand, Millipore, USA

4. METHODS

4.1. Cell Culture and Maintenance

Adherent cell lines (Melanoma Cell lines) used in this study were grown in DMEM - high glucose which contains the additives of 10% FBS, 1% penicillin/streptomycin and 1% non-essential amino acids. They were incubated in an incubator with the following conditions, 5% CO2 and 37°C. Passaging was done by removing the old medium and washing the cells with 1X PBS. Cells were detached by Trypsin (0.05%) in 37°C with a 3 minute incubation. After detachment was complete, equal amount of DMEM was added on the trypsinized cells to deactivate the trypsin activity, By pipetting, medium-trypsin and cells were mixed and resuspended. Proper ratio of cells were seeded back into new flasks or plates. Fresh medium were added in certain amounts.

In order to stock the cells for long periods; after cells were trypsinized and centrifuged, old medium was removed. Freezing medium added and cell pellet was resuspended in freezing medium 3.4. If 15-cm cell culture plate was fully confluent, it was divided into 8 cryovials which contains 1ml of cell and freezing medium mix. Cryovials were positioned into Nalgene Freezing Container filled with isopropanol. Container was kept at -80°C for a day or two and then cryovials were transferred into cryoboxes at -150°C. In order to thaw the cells from cell line stocks kept at -150°C freezers; one vial for one 75cm^2 flasks was thawed in two minutes in water bath. Immediately, after thawing was complete, they were transferred into 15ml falcon tubes previously filled with 3 ml of fresh and 37-degree-celsius DMEM. Cells were centrifuged at 2000rpm for 5 min. After aspirating the medium to remove DMSO, cells were resuspended and seeded into 15-cm plates.

Non-Adherent cell lines (DLBCL cell lines)used in this study were grown in IMDM which contains the additives of 20% FBS, 1% penicillin/streptomycin and 1% non-essential amino acids. They were incubated in an incubator with the following conditions, 5% CO2 and 37°C. Passaging was done by removing certain amounts of the old

medium and adding twice the amount of old medium to $75cm^2$ flasks.

In order to stock the cells for long periods; after cells centrifuged, old medium was removed. Freezing medium added and cell pellet was resuspended in freezing medium 3.4. If $75cm^2$ flasks was fully confluent (about 2x 10^6 cells per ml), it was divided into certain number of cryovials which contains 1ml of cell and freezing medium mix. Cryovials were positioned into Nalgene Freezing Container filled with isopropanol. Container was kept at -80°C for a day or two and then cryovials were transferred into cryoboxes at -150°C. In order to thaw the cells from cell line stocks kept at -150°C freezers; one vial for one $75cm^2$ flasks was thawed in two minutes in water bath. Immediately, after thawing was complete, they were transferred into 15ml falcon tubes previously filled with 3 ml of fresh and 37-degree-celsius IMDM. Cells were centrifuged at 2000rpm for 5 min. After aspirating the medium to remove DMSO, cells were resuspended and seeded into $75cm^2$ flasks.

4.2. Chromatin Immunoprecipitation (ChIP)

20-40 million cells were cross-linked with 1% formaldehyde on a shaking platform for 10 minutes at room temperature. Formaldehyde is a small molecule which passes through membrane and helps us with covalently binding IRF4 to its binding side and stabilizes it. Since excessive cross-linking reduces accessibility protein to its antibody and sonication efficiency and also it may mask epitopes, so glycine is added to quench the formaldehyde and cross-linking reaction was stopped by 0.125M glycine incubation for 5 minutes at room temperature. Glycine Cells were then washed two times in ice-cold PBS, resuspended in ice-cold ChIP lysis buffer supplemented with SDS and protease inhibitor cocktail, incubated on ice for 45min to 1 hour, and sonicated for 15 cycles (each for 1 min. at 80% duty cycle and 85% power output) with an MS72 tip-fitted Bandelin Sonoplus HD2070 sonicator. Sonicated lysates were cleared by centrifugation for 5 min at 13 krpm at 4°C. Soluble chromatin lysate derived from about 5 million cells was diluted three-fold with ChIP lysis buffer with protease inhibitor cocktail, and was incubated with 50μl of an equal volume mix of Protein G and Protein A magnetic beads pre-bound with 5μg anti-IRF4 antibody, or normal goat IgG for 45

min, at room temperature. Bead-bound immune complexes were then washed 6 times each for 1 min at room temperature with shaking using 0.8 ml of the following: ChIP Lysis Buffer (2 times), ChIP Lysis Buffer with 500mM NaCl (2 times), and TE (2 times). These washed, bead-bound immune complexes, and corresponding sonicated lysate (to be used as non-enriched 'input control') were then boiled for 10 min. in TE, treated with RNase A (0.1 μg/μl final concentration) for 45 minutes at 39°C (input controls only), treated with Proteinase K (200ng/μl final concentration) for 30 min. at 55°C, boiled again for 10 min, and the DNA residing in the supernatant was purified using silica-based spin columns.

4.3. Real Time Quantitative PCR (q-PCR)

Purified ChIP- DNA and input control DNA were diluted 25- to100-fold, and subjected to real-time PCR amplification in triplicates with region-specific primer pairs 3.2 on a PikoReal instrument (Thermo Scientific). qPCR reactions without template DNA were set up for each primer pair in order to rule out DNA contamination or unspecific amplification. The resulting qPCR data from each ChIP were then analyzed with the $\delta\delta$Ct method (taking into account the amplification efficiencies of the individual primer pairs as assessed by standard curves generated with DNA dilution series), normalized to corresponding input DNA samples' data, and plotted.

4.4. ChIP-seq

ChIP was performed as it is mention in section 4.2. Four ChIPs were pulled down together for each library preparation reaction. Library preparation is the essential step before sequencing. Chromatin immunoprecipitated-DNA fragments are end-repaired and adaptors ligated. In order to multiplex different samples in a single lane of sequencing machine, adapters are indexed with specific barcodes which allows user to identify in silico the source of the read during analysis step. At the end of library preparation, products are size-selected and purified. Final step is amplification with PCR [81, 85]. In this project Library preparation was done with NebNext Ultra DNA library preparation as mentioned by manufacturer and then Library prepared samples were dried in

DNAstable tubes according to manufacturer's protocol. During the sequencing, in Illumina sequencing platform, adapter ligated fragments are attached to the flow cell with the help of single stranded oligonucleotides bound to the surface. Oligos are complementary to adapter sequences. Attached DNA fragments are "bridged amplified" just before sequencing. During bridge amplification, firstly, priming occurs at the free end of the fragment to the complementary sequence on the surface. This results in a shape like a bridge. Polymerase-based amplification with unlabeled dNTPs produce millions of clusters. Since imaging systems can't detect single fluorescent events yet, cluster generation is essential. Those clusters are sequenced along with synthesis after the addition of all four labeled and reversible terminators, primers, and DNA polymerase. At each cycle, a single fluorescent nucleotide incorporates to the newly synthesized complementary fragment. After each single base extension, emitted light is detected by high resolution imaging system and corresponding nucleotide is determined [76].

4.5. Bioinformatic Methods for ChIP-seq Analysis

4.5.1. Data Quality Control by FastQC via Galaxy Web Server

The raw data which researcher starts bioinformatics analysis is in "fastq" format. Here, depending on their barcodes each read is grouped, so reads originated from different samples are de-multiplexed which means bioinformatically they are separated. After grouping the reads, adaptor/index sequences should be removed. In the fastq format, each read has its four informative lines which contain a record identifier, the sequence, and also quality scores corresponding to each base. Those scores are Phred quality values assigned depending on base-calling accuracy. According to quality scores, low scored bases can be trimmed for a better alignment to the genome.

FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was installed. The fastq files were analyzed with default parameters in the FASTQC program.

4.5.2. Fetching Sequenced Total Read Number From a FastQ File

All of the reads sequenced for each sample are stored in separate fastq files. Command line to learn how many reads there are in a fastq file is shown below:

```
cat <raw data> .fastq mid paste - - - - \mid wc -L
```

4.5.3. Aligning Reads to the Human Genome by Bowtie Software

In a standard ChIP-seq bioinformatic analysis pipeline, aligning (mapping) is the next step after quality control. Parameters, like mismatch number or mapping uniqueness, can also be adjusted. Bowtie is the most common used program for alignment of the reads to genome [86].

Bowtie Software v1.1.1 (http://bowtie-bio.sourceforge.net/index.shtml) was installed to the local server of the department, which runs on Ubuntu (12.04.1 LTS) operating system. It was run remotely via PuTTY software (http://www.chiark.greenend.org.uk/ sgtatham/putty/download.html). Command line to run the program with selected parameters (http://bowtie-bio.sourceforge.net/manual.shtml) in Ubuntu, is given below:

```
./bowtie -n 3 {L 45  -m 1 -5 3 -3 2 --best --strata hg19
<sample-name.fastq> <out-putfile.bam>
```

./bowtie gives the path to the bowtie program and tells Ubuntu to run it.

-n 3 –l 45 tells Bowtie to accept no more than 3 mismatch between a the 45 bp of the sequence read and its best homologue in the genome.

-m 1 tells Bowtie to reject any reads that are identical to more than 1 sequence in the genome (since it would be unknown which locus our read really came from).

-5 5 tells Bowtie to trim the first 3 (lower quality) bases from the read before mapping.

-3 2 tells Bowtie to trim the last 23 (lower quality) bases from the read before mapping.

–best and –strata tell Bowtie to try hard to find the best match..

Human genome sequence index, which is the reference genome file, was downloaded from Illumina's iGenomes collection with hg19 build.

4.5.4. Mapping Statistics by SAMtools

SAMtools is a program to acquire information from or perform manipulations on the files in bam or sam format [39]. Software was installed to the local server. Input files were the "sample-name.bam" (filtered for uniquely mapped reads) files. Command line to learn the read number in a bam file is given below:

```
samtools flagstat <name.bam>
```

4.5.5. Merge Aligned Reads From Different Lanes

Each of our sample data came in three parts. So after aligning them to reference genome, in order to continue the analysis, we need to merge all bam files of each sample together. This is done by the following command in SAMtools:

```
samtools merge <out.bam> <in1.bam> <in2.bam> <in3.bam>
```

out.bam is the output file's name and then we write the name of the input files.

4.5.6. ChIP-seq Qulaity Control

The standards and quality measures for a perfect ChIP-seq is still a hot topic in the field of epigenetics and NGS sequencing. Difference between antibodies, biological replicates, sequencing depth and immunoprecipitation (IP) are just a few parameters which effect ChIP-seq experiments and data. Recently several tools on different programming platforms have been developed in order to evaluate the quality of ChIP-seq experiments. NGS-QC Generator is one of them. The remarkable side of this program is, they have also a large database of quality control reports from publicly available data in particular ENCODE datasets where you can compare your quality control reports to to ENCODE-project-related report. NGS-QC generates read count intensity (RCI) profiles of randomly selected subsets of the total mapped reads (TMRs) of ChIP-seq data and then defines the deviation from the theoretical expected read count intensities. So, TMRs are randomly sampled at three different densities (90%, 70% and 50%); and then their genomic RCI profile compared to that of the original profile. Then the program as-signs a three-letter grade between the AAA-DDD. For transcription factor ChIP-seq samples the AAA is the perfect score but in the database, the grade CCC is still satisfactory and acceptable. However for the negative control or normalization sample, DDD is the perfect score which means there is no selectivity (bias) over the genome, unlike the ChIP-seq samples. However in the NGS-QC database in the section related to quality reports of input DNA, there are a lot of quality reports where the score is close to AAA suggesting that there is a bias in the input and some over-amplification in selected regions of genome, which is not desired for input [87].

4.5.7. Subtracting UCSC Blacklist From All Reads

BEDtools utilities are tools for a wide-range of genomics analysis and manipulation tasks. For instance, With BEDtools we can intersect, merge, count, complement, subtract and shuffle genomic intervals from multiple files in widely-used genomic file formats such as BAM, BED. All different analysis by this software are conducted by combining multiple bedtools commands on the Ubuntu command line. The following command helps us with removing precentromic and telomeric regions from our

sequences. These region are presented as a bed file in UCSC genome browser.

```
bedtools subtract -a <sample name.bed> -b <UCSC-blacklist-hg19.bed>
```

-a is our sample file which the operation will be carried on it. -b the region which will be subtracted is indicated in this file.

4.5.8. Peak Calling (MACS, Homer, ByesPeak)

Following to mapping, peak calling should be performed in order to find possible TF binding regions in the genome. In brief, peak callers compare number of reads in chromatin-immunoprecipitated-DNA samples for different genomic regions to number of reads in input samples in those genomic regions. Hence, when peak caller assigns a peak to a region, it means compared to input, there are more reads in that region in ChIP samples. We used three programs for peak calling: MACS, Homer and BayesPeak. MACS is one of the most renowned program in the field of ChIP-seq peak calling which is supported and updated frequently [88]. Homer and BayesPeak are one of the first program packages which provide user with a pipeline. Homer is based on Perl and BayesPeak is based on R [89–91].

Model-based analysis of ChIP-seq (MACS) detects genome-wide locations of transcription/chromatin-binding factor or histone modifications from ChIP-seq data. For peak calling MACS follows four steps: redundant reads removal, adjusting read position (extension of reads), calculating peak enrichment (compared to background) and estimating the empirical false discovery rate (FDR). MACS identifies positive peaks then swaps control and treatment data (ChIP samples) in order to identify negative peaks [88].

In Homer, the identification of peaks was performed by finding significant clusters of tags in ChIP-seq samples. The next step is filtering these clusters for the significantly enriched compared to background and local ChIP-seq signal within sliding window of

200 bp. Putative peaks are searched for based on the 4 principles: 1) The number of tags in each cluster must surpass a threshold corresponding to a false discovery rate (FDR) of 0.1%. 2) Neighboring peaks must be greater than 500 bp away from one another. 3) Peaks must have at least 4-fold more tags (normalized to total tag count) than the input control sample from the same cell type. 4) Peaks must have 4-fold more tags per bp in the peak region (200 bp) relative to the surrounding region (10 kb) to avoid identifying regions with genomic duplications or peaks without localized binding [89].

The third peak caller BayesPeak, which is part of Bioconductor package, identify the peaks based on Bayesian hidden Markov model. BayesPeak have the advantage of allowance for overdispersion in read counts and a competitive genome-wide specificity and sensitivity. By identification based on peak structure, BayesPeak does not just rely on peaks based on total numbers of reads in that region, but also checks them for appropriate structural formation of the reads inside a possible peak. Another important feature of the algorithm is applying the negative binomial distribution to model the counts of sequenced reads. This allows for overdispersion and provides a better fit to the data than the Poisson distribution which is widely been used by various methods [90, 91].

For analysis with different peak callers, depending on the software, the language of command lines were different. For MACS it was python, for Homer it was perl and for Bioconductor we used R. For MACS, the command for peak calling is:

```
Macs14 -t <ChIP-sample.bam> -c <input-Sample.bam> --format=BAM  -m 7,12
--wig {S -B  -S --slocal=1000 --name output file name
```

-t is used for indicating the name of the ChIP sample

-c is used for the name of input sample which will be used for normalization

-m indicate that for calculating peak model from top 1000 peaks, MACS should search for peaks which were minimum 7 times more enriched than the background (input file) and their maximum enrichment should not pass over 12 folds.

–wig –S tells MACS to save enrichment profiles for both files in wig format in a single file

-B -S tells MACS to save enrichment profiles for both files in bedgraph format in a single file

–slocal=1000 The small nearby region in base pairs to calculate dynamic lambda. This is used to capture the bias near the peak summit region. Normally, MACS calculates a dynamic local lambda to reflect the local bias due to varying enrichment in the input sequences.

–name specifies the name of the output files

For Homer,first we need to prepare tag directories which basically means to analyse each alignment file and splits them into separate files based on their chromosome number.

```
makeTagDirectory <output file name> <sample name.bed> -format bed
```

-format bed specifies that the input files are in BED format

As a result of this command, several *.tags.tsv files are created in a directory with the name of assigned output name. This helps speed up the analysis of large sequencing data without running out of memory.

Then for peak calling we use the following operation:

```
findPeaks $<tagDirectory of each replicate> -style factor -o
$output file name> -i <input tag directory>
```

-style factor specifies that ChIP-seq experiment was for a transcription factor

-o assigns and output file name

-i is for indicating the tag directory for input file.

In BayesPeak, after installing the required packages, the following command in R is executed. First we loaded the program and its packages:

```
library $(BayesPeak)$
```

The following command will find possible peaks in the sample:

```
raw.output <- bayespeak ("ChIP-sample.bed","Input-sample.bed")
```

Then, you summarize and report the peaks with the following command:

```
output <- summarize.peaks (raw.output, method = "lowerbound")
```

raw.output is a list - it contains not only the bins called, but also some useful QC information (such as the model fit).

summarize.peaks() is used to summarize the raw.output object. This combines the raw bin calls into peaks and combines data across jobs.

4.5.9. Read to Peak Ratio Calculation and Graph

For analyzing read to peak ratio and saturation status of the ChIP sample with higher read density, random sampling was performed for various number of reads (5% of all reads, 10% of all reads, ... , 90% of all reads). We calculate number of peaks for each randomly-sampled read collection by MACS. For each percentage ratio, we took 10 random samples from the the complete dataset of reads.

4.5.10. Finding Peaks Common Between Both Samples and Algorithms

For finding the overlap of peaks primarily across the replicates for each peak calling program and then we will get the overlap of the overlapped-peaks of the replicates. Consequently, the final file will contain overlap of the replicates and peak-calling programs. For this purpose we will use the following command from BEDtools suite several times:

```
intersectBed -a File-A.bed -b File-B.bed $>$ common peaks.bed
```

4.5.11. Get FASTA Sequences From Bed Files

Input file for motif analysis should be in FASTA format. In order to get fasta sequences of the peaks, the following command from BEDtools were used:

```
bedtools getfasta  -fi hg19.fa -bed commonpeaks.bed {fo
commonpeaks.fa.out
```

-fi hg19.fa is the fasta file of human reference genome. This file is used for finding the sequences

-bed indicates the bed file which here is the common peaks file

-fo refers to the name of output file which will be in FASTA format.

4.5.12. Motif Analysis by MEME-ChIP

MEME-ChIP is web-based program designed for discovering motifs in large sets of short (around 500bp) DNA sequences which are produced by ChIP-seq experiments. We give MEME-ChIP our sequences from peaks in FASTA format. MEME-ChIP performs a five-step job. First, it performs *ab initio* motif discovery, then motif enrichment analysis. Next, it visualizes the motifs and analyze the binding affinities and finally identify the motifs. MEME-ChIP applies combination of motif discovery using MEME and DREME and also compares both found motifs and the sequence data against databases of known motifs [92, 93].

4.5.13. GREAT: Genomic Regions Enrichment of Annotations Tool

After peak determination, peaks need to be assigned to nearest genes. After finding the target genes, gene set enrichment programs to get an insight into the biological meaning or pathways related with those assigned enriched genes.

GREAT (http://bejerano.stanford.edu/great/public/html/index.php) gives biological meaning to a set of non-coding genomic regions by analyzing the annotations of the nearby genes. Therefore, it is mainly useful in studying cis functions of sets of non-coding genomic regions. Cis-regulatory regions can be identified via experimental methods such as ChIP-seq. After identification of peaks and transcription factor binding site, we uploaded the list of common peaks across both replicates and three peak caller programs to GREAT website. GREAT finds enriched annotations among the genes which are near to the binding sites and then lists the possible processes which are regulated by our desired transcription factor [94].

4.5.14. Venn Diagrams

Venn Diagrams were created by "Venn Diagram Plotter" (http://omics.pnl.gov/software/venn-diagram-plotter). In order to visualize common genes between ChIP-seq data and IRF4 differentially regulated genes (RNA-seq data), we uploaded the list of genes in .txt format and visualized them with Venn Diagram Plotter.

4.5.15. GOrilla and REVIGO

For identifying gene ontology terms for the common genes between RNA-seq and ChIP-seq data, we used GOrilla (Gene ontology enrichment analysis and visualization tool) and REVIGO for analysis and visualization respectively. GOrilla is an resourceful Gene Ontology analysis tool with unique features that make a useful addition to the existing collection of Gene Ontology enrichment tools. GOrilla is publicly available (http://cbl-gorilla.cs.technion.ac.il). One of the options of GOrilla is to export the data to REViGO ,a web server program for better visualization of GO terms. REVIGO takes long lists of Gene Ontology terms and remove redundant GO terms then remaining terms are visualized in semantic similarity-based scatterplots, interactive graphs, or tree maps [95, 96].

5. RESULTS

5.1. Optimal Average Fragment Size for Chromatin Immunoprecipitation

Average sheared chromatin size during ChIP sample preparation is a critical parameter. There are various methods to shear chromatin, whether mechanical or enzymatically (with Micrococcal Nuclease). Here, we utilize and optimize the sonication process in order to achieve mostly 200-700bp average DNA size. Figure 5.1shows average size of sheared chromatin under 3 different sonication conditions in SkMel-28 cell line, melanoma cell line. All cellular lysates are from SkMel-28 cell line. Sonication conditions were the same among all samples, the only difference between them is the number of sonication cycles. With increasing number of sonication cycles, smaller chromatin fragment size can be achieved. In Figure 5.1, each lane presents different number of cycles of sonication. Lane C and D sonication results are in our desired average size range. However since the average fragment size in the middle of the smear is lower than 500bp, the number of sonication cycles were set to 15 cycles. There is high possibility of losing most of protein-DNA cross-linked region due to too many cycles of sonication that will lead to over-shearing which is not desirable in ChIP experiment.

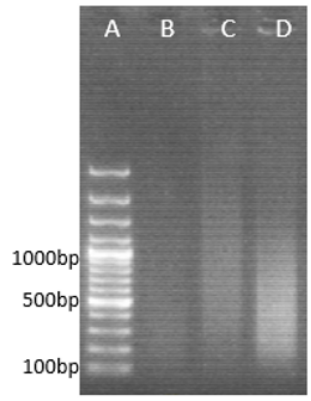


Figure 5.1. Agarose gel electrophoresis of DNA sonication optimization. Lane A is the DNA ladder molecular weight marker. Lane B, C and D are sonicated purified DNA from SkMel-28, a melanoma cell line which were sonicated 9-13-17 cycles respectively (Each cycle is 1 minute with 85% power and 90% duty cycle).

5.2. Establishment and Optimization of ChIP Assays in Melanoma Cell Lines

After sonication optimization, the next step is setting up and performing anti-IRF4 ChIP in melanoma cell lines. In order to have functional straightforward ChIP protocol, we needed to optimize reagents and protocols such as: designing and testing different PCR primers from binding regions in the genome, choosing the best antibody, and trying different buffer conditions for washing the magnetic immunocomplex.

5.2.1. Primer Optimization

In order to implement anti-IRF4 ChIP experiments in melanoma cell lines, we turned to previously known top IRF4-binding regions in multiple myeloma and ABC-like diffuse B-cell lymphoma as candidates for binding also in melanoma. Therefore, the primers were designed from previously identified IRF4-binding regions in ABC-DLBCL, one of which belongs to an intronic region of EHBP1L1 gene. For this regions, strong peaks in ChIP-seq data sets were observed in multiple myeloma and ABC-DLBCL.

Figure5.2, qPCR for designed primers from one putative IRF4 binding region is shown. The designed primer set belongs to possible IRF4 binding region on EHBP1L1 region. As Figure 5.2 demonstrates, the EHBP1L1-2 showed both good enrichment and good melting curve is so from now on, we will continue with the mentioned primer.

5.2.2. Selection of the most suitable IRF4 antibody for ChIP

For the determination of the most suitable anti-IRF4 antibody, we conducted ChIP with two different anti-IRF4 antibodies: one polyclonal IRF4 antibody (Santa Cruz) which is used in ENCODE project; and another IRF4 antibody which in monoclonal (Cell Signalling). NLRP7 antibody were used as negative control antibody because NLRP7 protein is a cytoplasmic protein with no known target in nucleus. Normal Rabbit IgG were used as negative control for IRF4 antibody from cell signaling. The ChIP-qPCR results from Figure 5.3 indicates that both antibodies are

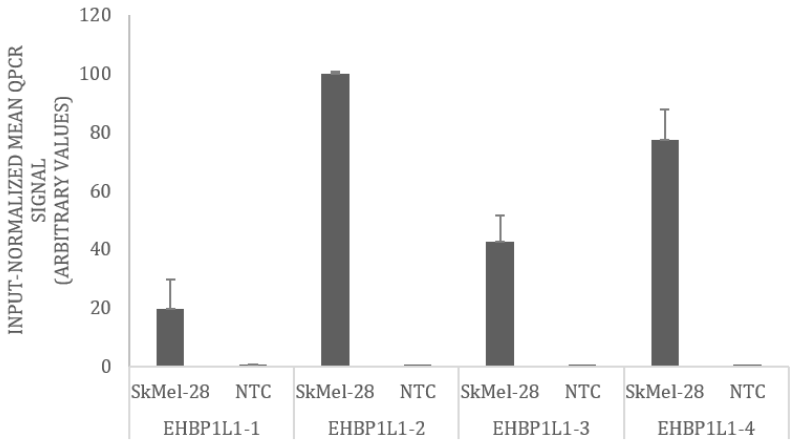


Figure 5.2. Primer Optimization for IRF4 binding region, EHBP1L1. qPCR was done for each of EHBP1L1 primers. Template is purified sheared DNA from SkMel-28. NTC stands for no template control.

functional.

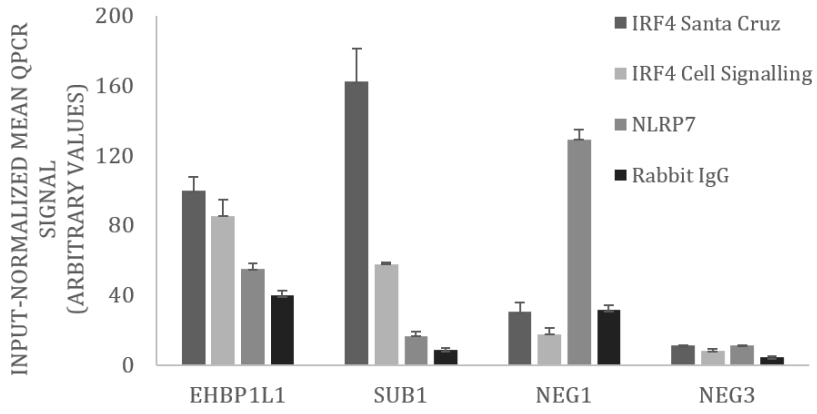


Figure 5.3. ChIP-qPCR for IRF4 antibody from different companies. EHBP1L1 and SUB1 are designed from regions with possible IRF4 binding. NLRP7 and Rabbit IgG as mentioned before, are negative control antibodies. NEG1 and NEG3, the negative control primers are genomic region with no apparent IRF4 binding. Error bars depict standard deviation of the mean.

Based on the experiments to find the most efficient conditions for ChIP-qPCR and ChIP-seq we tried different immune-complex washing solution during our ChIP-qPCR experiments (A.1). As the results from this section showed, with the right choice of antibody and immune complex washing routine (i.e., with a wash solution without SDS detergent), can improve the ChIP-qPCR results.

5.2.3. Validation of ChIP-qPCR in Melanoma Cell lines

The next step is to validate ChIP-qPCR in melanoma cell lines. Therefore, we performed anti-IRF4 ChIP-qpCR in two different melanoma cell lines, SkMel-28 and SkMel-5 and two different DLBCL cell lines, HBL-1 and OCI-LY-19. HBL-1 is an ABC-DLBCL cell line with high IRF4 expression. HBL-1 will be our positive control cell line. OCI-LY-19 is from GCB-DLBCL family of cell lines with no IRF4 expression and is used as our negative control cell line.

Our results, Figure 5.4, verified that anti-IRF4 ChIP in melanoma cell lines is working with relatively high amplification signal at putative IRF4 binding region which

is EHBP and SUB1 which is previously identified IRF4-binding region from ABC-DLBCLs.

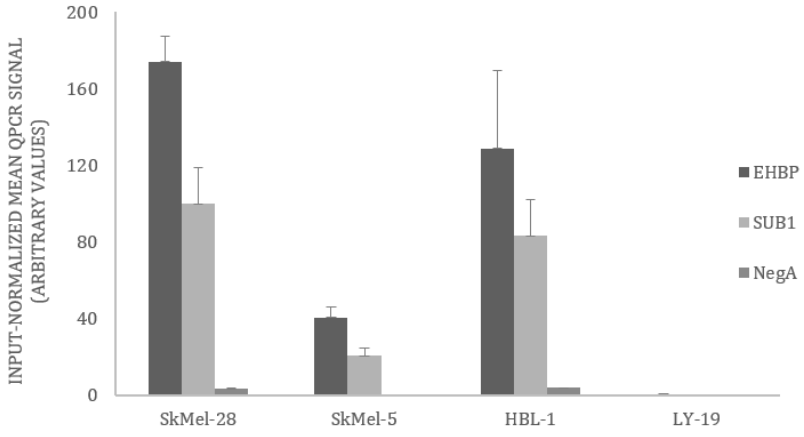


Figure 5.4. ChIP-qPCR with IRF4 antibody in Melanoma and Lymphoma cell lines. EHBP1L1 and SUB1 are designed from regions with possible IRF4 binding. NegA, the negative control primers is a genomic region with no apparent IRF4 binding. Error bars depict standard deviation of the mean.

5.3. Identification of IRF4 binding region on Tyrosinase Promoter in Melanoma Cell Lines

Tyrosinase (TYR) is a gene encoding pigmentation enzyme in melanocytes. TYR is essential for development and function of melanocytes [97]. Previous studies have shown elevated expression of melanocytic lineage genes in melanomas [98]. Our aim was to show whether IRF4 has binding regions on TYR promoter. First, we identified putative IRF4-binding regions on TYR proximal and distal promoter regions via analysing DNase-seq data from the ENCODE project and found possible IRF4 binding motifs (5′-GAA(A)-3′) identified previously [43]. As it is demonstrated in Figure 5.5, we identified two regions, one close to transcriptional start site (TSS) on proximal promoter and another region more distal, 2kb upstream of TSS . We designed primers from these two putative regions and performed ChIP-qPCR in SkMel-28, SkMel-5 and OCI-LY-19 cell lines.

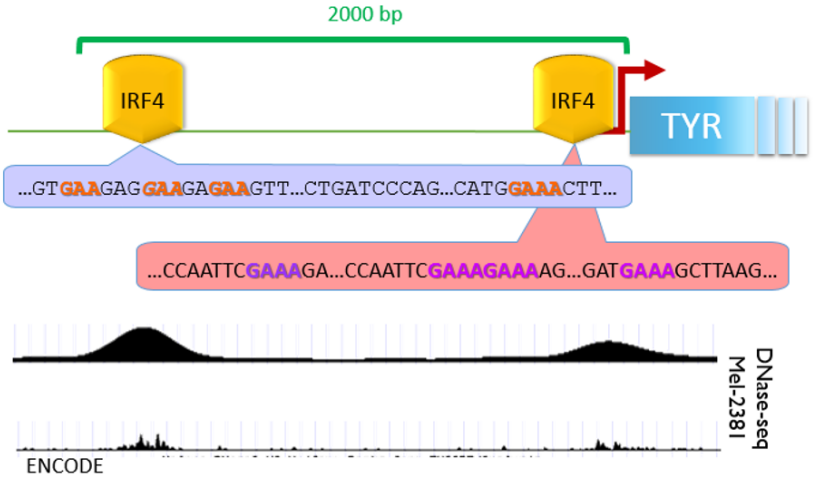


Figure 5.5. Putative IRF4 binding sites on schematic TYR locus structure. Highlighted bold sequences are possible IRF4 binding motifs. DNase I hypersensitivity mapping on TYR promoter region in Mel-2381 cell line from ENCODE project from UCSC Genome Browser.

The results in Figure 5.6, shows that IRF4 binds to proximal promoter of TYR as well as the distal promoter region in SkMel-28 and SkMel-5 cell lines.

5.4. Sample preparation for ChIP-seq

After establishing and optimization of ChIP assay in melanoma cells, we performed ChIP-seq and identify genome-wide targets of IRF4 in melanoma cell lines. The experimental part starts with culturing the cells, and prepared lysate from $50x10^6$ cells and sonication. Figure 5.7 shows average fragment size from sonication of samples for ChIP-seq which is mostly in the 200-700bp range.

Then we performed four ChIP experiments for each replicate, pulled them down together to prepare ChIP-seq samples. Verification of ChIP results were done by qPCR. After verification, library preparation for next generation sequencing were performed. End-repair was done and then adapters were ligated to the end of fragments (with different index barcodes for multiplexing). The size selection step was performed for the size range between 200-700bp adaptor-ligated fragments. After size selection, there was final round of amplification via PCR. Samples were shipped to and sequenced at

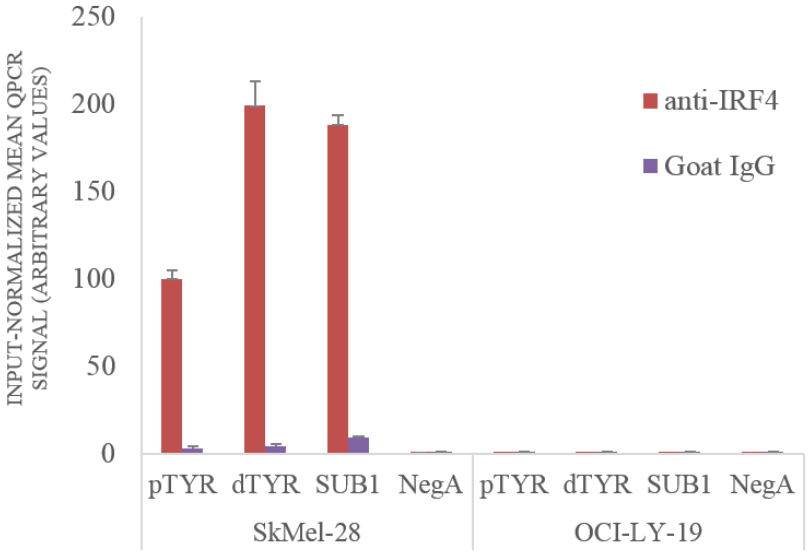


Figure 5.6. Quantitative anti-IRF4 ChIP for TYR promoter. SkMel-28 and OCI-LY-19 cell line. The TYR primers are indicated as pTYR for proximal promoter region and dTYR for distal promoter region. SUB1 is the positive control primer pair and NegA is the negative control primer pair.

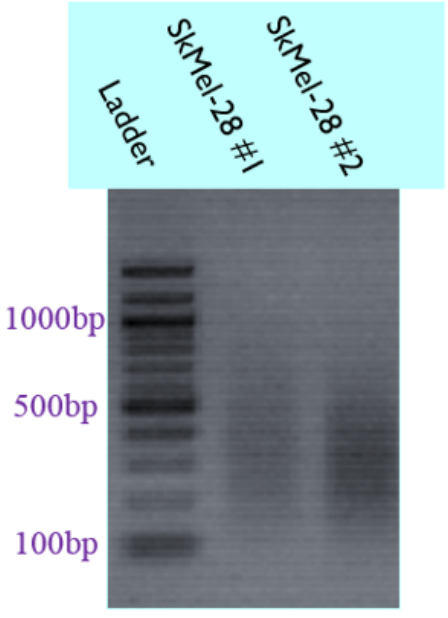


Figure 5.7. Agarose Gel Electrophoresis of Sonicated two biological replicates of SkMel-28. After sonication the samples were run on a 2% agarose gel.

University of Texas at Austin, Genomic Sequencing and Analysis Facility (UT GSAF). Before sending we did another round of qPCR, in order to compare with the results of before-library-preparation qPCR. The results are shown in Figure 5.8.

According to Figure 5.8, enrichment profile of the replicate sample for before and after library preparation shows about 25-30 times increase as expected. So for the last quality control step, we loaded small amount of samples in Agilent Bioanalyzer to check the integrity and concentration of the samples. The average size and concentration of the samples are shown in C.1 More than 90% of the fragments in all samples are around the expected average size. Also the concentration of the samples are enough for next generation sequencing as designated by UT GSAF facility.

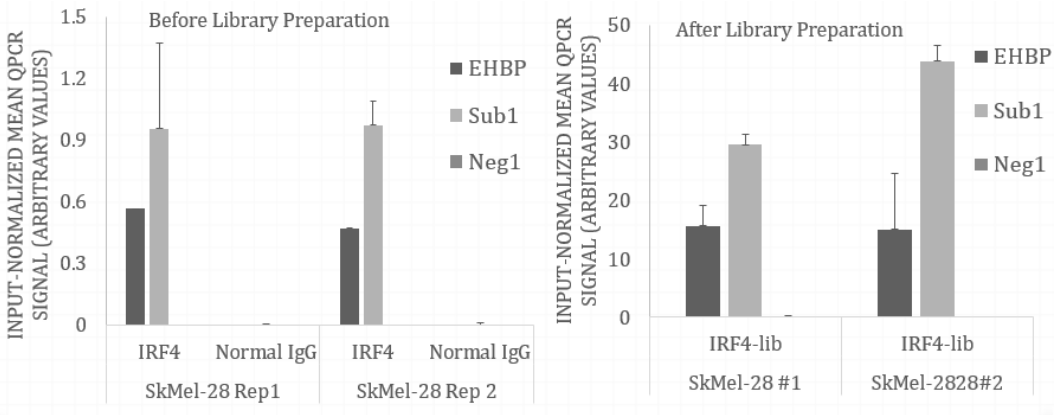


Figure 5.8. ChIP-qPCR verification of ChIP-seq samples,before and after library preparation. EHBP and SUB1 are positive control primers from the regions where IRF4 is possibly binding; and Neg1 is negative control primer with no evidence of IRF4 binding in that region.

5.5. Bioinformatical Analysis of ChIP-seq Results

UT GSAF Sequencing facility uploaded the sequenced reads on their webserver in fastq format 4.5.1. They have already de-multiplexed and removed the adaptor sequences along with barcodes. We set up and followed the bioinformatical analysis pipeline summarized in Figure 5.9 and Section 4.5

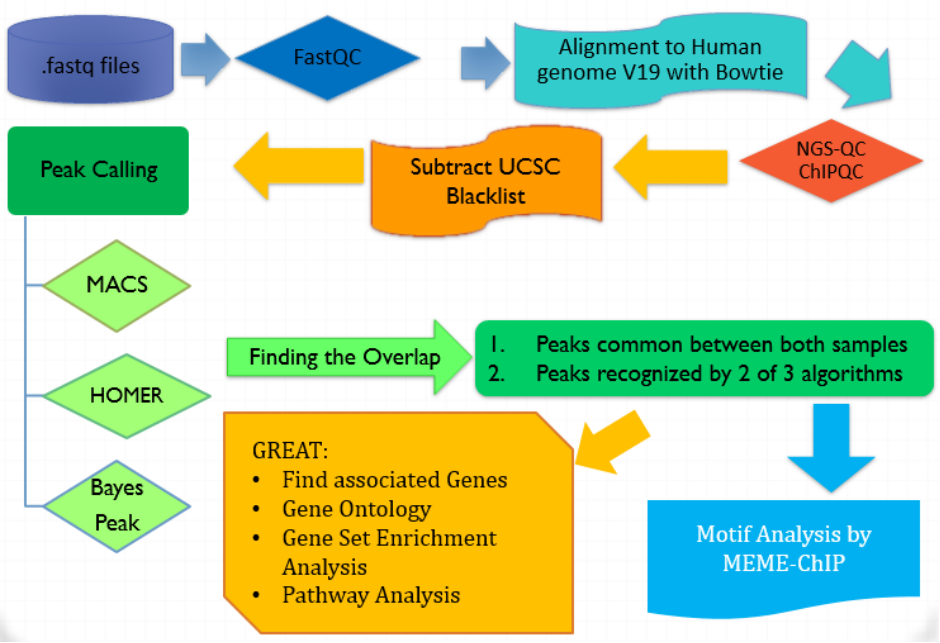


Figure 5.9. Schematics of followed bioinformatical ChIP-seq analysis pipeline.

5.5.1. Quality Check for Sequenced Reads with FastQC

First step of every NGS analysis pipeline is analyzing quality of the reads by FASTQC. We can use the program both through Galaxy web server or download it and run locally. FASTQC helps us to decide if any trimming or filtering is necessary. As an illustrative figure, quality score graph for one of the biological replicates SkMel-28 sample is shown in Figure 5.10. All of the quality graphs of each biological replicate are shown in Figure D.1.

Before alignment, 3 bases were trimmed from $5'$ side and 2 bases from $3'$ side in order to avoid any possibility of low quality base calls at the beginning and ending of the reads. There was no need for filtering since qualities of bases in each position were satisfactory (most of the bases have a Phred quality score more than 34 (base call accuracy $> 99.9\%$)). Hence, all the reads were used at the mapping step.

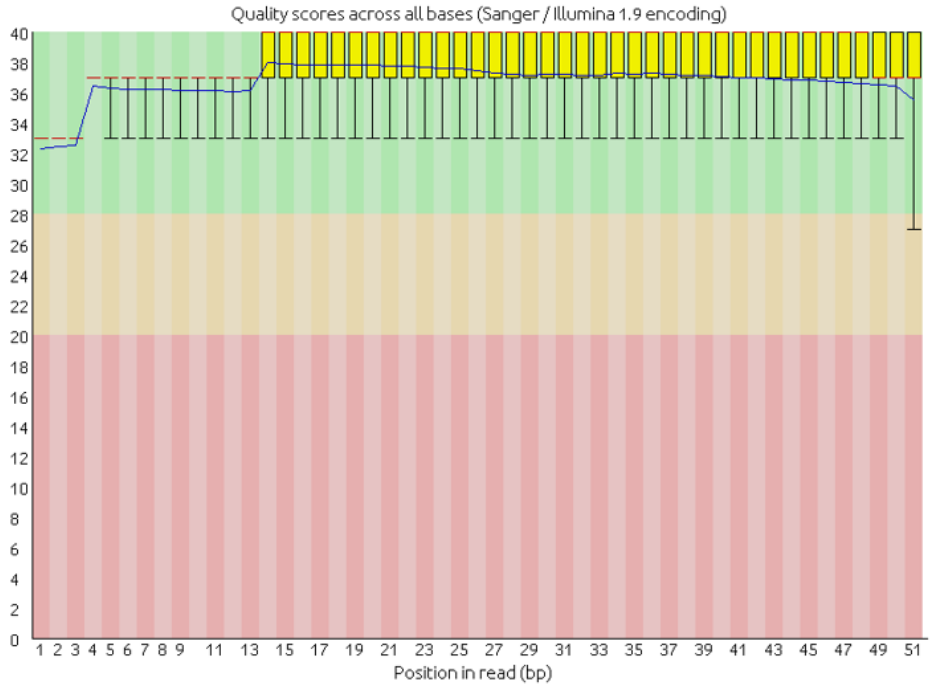


Figure 5.10. Per base sequence quality graph for one of the biological replicates of SKMEL-28. X-axis shows base position and Y-axis point to a Phred quality score.

5.5.2. Alignment to Human Reference Genome (Hg19) with Bowtie

We used Bowtie [86] to align sequenced reads to human ge-nome (hg19). In order to avoid aligning to more than one region of genome, we used the option of "unique mapping", which are the reads mapping only to one region of the genome, as explained in Materials and Methods section 4.5. Mapping statistics is presented in Table 5.1.

Table 5.1. Bowtie Mapping Statistics.

SkMel-28	**Total**	**Uniquely Mapped**	**Mapped %**
ChIP-IRF4 Replicate-1	50,644,130	33,286,423	65.73%
ChIP-IRF4 Replicate-2	88,679,913	56,570,355	63.79%
Input	97,574,614	75,801,925	77.69%

We visualized the reads aligned by Bowtie using the UCSC Genome Browser. A representative figure, showing MITF locus for all replicates and input of SkMel-28, is shown in Figure E.1.

5.5.3. NGS-QC: A Quality Control Tool for ChIP-seq Data

The quality report for SkMel-28 samples and input can be seen in F.1. Assigned grades for our samples were as follows: SkMel28-IRF4-Replicate 1 (CBB), SkMel28-IRF4-Replicate 2 (BAA) and SkMel-28 input (BAA). So because of the quality score of input which shows bias and over amplification in some selected regions, we have decided to perform peak calling with more than one peak caller and continue the pipeline based on the overlap of different algorithms and common peaks between both samples.

5.5.4. Peak Calling

Peak calling and identifying the significant peaks has always been one of the challenging steps in ChIP-seq analysis pipeline. In order to increase fidelity and overcome the problems originated from input (as mentioned before briefly in the last section), we decided to identify peaks by applying the following three programs: MACS, Homer, and BayesPeak. The common peak calling criteria for all three programs were chosen to be FDR equal or lower than 0.1 which means on the average there is only 10% chance that the defined peak is false positive.

In MACS, one of the clues that shows ChIP-seq has worked is the ratio of positive peaks to negative peaks should be around 10:1. Statistics of Peak calling by MACS is shown in Table 5.2 and modeled peak based on top 1000 peaks is demonstrated in G.1.

Table 5.2. Peak Calling statistics of MACS.

Samples	**Positive Peaks**	**Negative Peaks**	**Common Peak between replicates(R1∩R2)**
Replicate 1 vs Input	166,301	13,224	82,989
Replicate 2 vs Input	297,942	41,258	

As our second peak caller, we used Homer [89]. According to the aforementioned four criteria in Section 4.5, identified peaks for SkMel-28 Replicate 1 is 121,154 and for SkMel-28 Replicate 2 is 126,458 peaks. The common peaks between two samples is 26,995 peaks.

The common peaks between two replicates in MACS analysis is 49.90% and in analysis by Homer it is about 22.28%. This difference may be because of the fact that number of reads in replicate 1 is almost half of Replicate 2. However, in order to confirm significance of the peaks we used a third computational algorithm for peak calling that belongs to Bioconductor packages.

Additionally, we decided to check the peak saturation levels in the ChIP sample with higher number of reads. In brief, we randomly sampled from ChIP-seq replicate 2 (5%-10%-15%...-90% of the reads). We performed this random sampling 10 times for each percentile. Basically it means that each random sampling was performed 10 times in order to overcome the effect of variation across the genome and reads. Peak calling was performed by MACS on all of the randomly sampled subsets and later on, the graph was drawn based on the number of peaks. As Figure 5.11 suggests in order to achieve saturation and find all IRF4 binding sites, we need more reads.

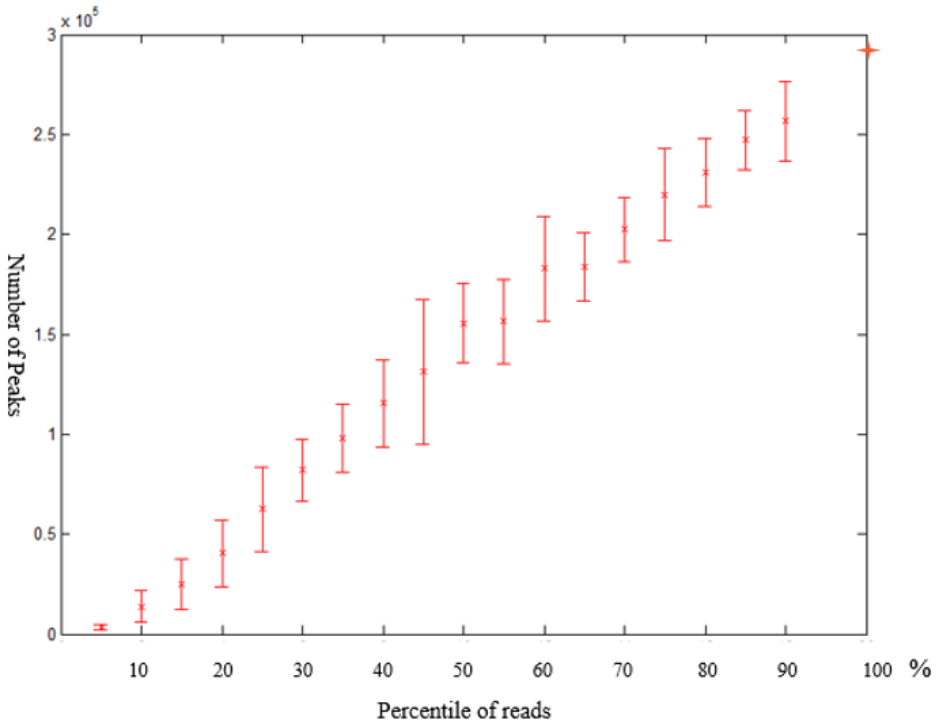


Figure 5.11. Variation of number of peaks based on random sampling of peaks. Less number of reads leads to less number of peaks. More reads suggests more IRF4-binding sites are discovered.

The third peak caller, BayesPeak which is part of Bioconductor package identify the peaks based on Bayesian hidden Markov model [90]. With FDR of 10%, BayesPeak discovered: 102,215 peaks for SkMel28 Replicate 1 and 108,697 peaks for SkMel-28 Replicate 2, and the number of common peaks between both samples are 29,735 peaks. Table 5.3 is summary of number of peaks across different platforms.

Table 5.3. Summary of Peak Caller Statistics.

All Peaks	**Replicate 1**	**Replicate 2**	**Common Peak between replicates (R1∩R2)**
MACS	166,301	297,942	82,989
Homer	121,154	126,458	26,995
BayesPeak	102,215	108,697	29,735

In order to be more stringent in downstream analysis, we decided to choose the peaks which overlap in two categories: first they are present in both replicates and second, across different peak callers, at least 2 of 3 algorithms have discovered it. The common peaks pool is 23,101 peaks. These are the peaks which most of downstream analysis will be based on.

5.5.5. Motif Analysis using MEME

For finding motifs and their enrichment pattern over the peaks, we used MEME-ChIP of The MEME Suite. We give MEME-ChIP our sequences from peaks in FASTA format [92, 93].

Figure 5.12 shows enriched motifs in common peaks dataset of the different algorithms as were explained and selected in Section 5.5.4. In Figure 5.12, IRF motif (red) is not enriched in center of the peak defined by MEME-ChIP, but about 20-40bp away from center of the peaks, enriched IRF4 motif can be seen. As it has shown in Figure 5.12, IRF motif is grouped with NFκB motif (yellow), TFAP2A motif (violet), MEF2A (pink) and MEF2B (light green). Centrally enriched motif 5′-AAAA(T/A)TA(G/C)-3′, is an AT rich which is similar with motifs from NFAT5 and MEF2A and MEF2B. The related motif was ranked first with highest motif significance value which is calcu-

lated based on p-value and the number of candidate motifs tested.The relate p-values can be seen in Figure 5.12.

Additionally, we decided to check for enriched motifs in top 25% common peaks between both replicates from each peak caller separately. In Table H.1, enriched motifs with highest score from MACS's top 25% peaks are shown. In this dataset, enriched motifs from IRF, TFAP (Transcription factor AP/activating enhancer binding protein), NFAT (Nuclear factor of activated T-cells),and ZNF (Zinc finger proteins) transcription factor families can be seen. Table I.1, demonstrates highest-scored enriched motifs for top 25% of peaks called by HOMER. We can see transcription factors (TF) such FLI1,a member of ETS family. Some of ETS family members are known IRF4 cofactors in ABC-DLBCLs. Other enriched motifs belong to TFs such as IRF, MEF2 (myocyte enhancer factor-2) and ZNF.

Top 25% peaks of BayesPeak related enriched motif results were different. As you can see in Table J.1,we can see partially enriched IRF enriched motif in the form of 5′-GAA(A)-3′, however IRF motif is not in top 10 enriched motifs in BayesPeak dataset.

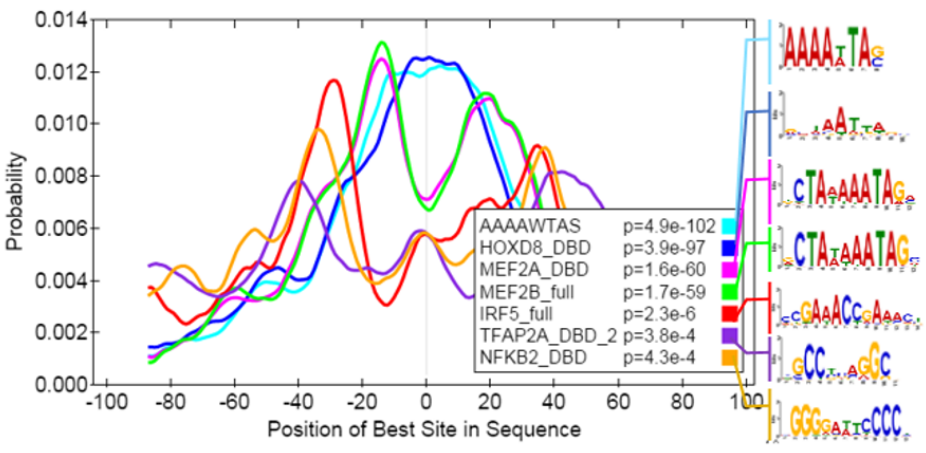


Figure 5.12. Various Transcription factors motif enrichment pattern in the common peaks. Graph and sequence logo are generated by MEME Suite.

5.6. Linking Peaks to Genes, Gene Set Enrichment Analysis, Pathway Analysis using GREAT

Genomic Regions Enrichment of Annotations Tool (GREAT), a web-based program, assigns biological meaning for a set of known or expected cis-acting non-coding genomic regions by analysis of the annotations of the nearby genes. One of the main experiments which GREAT is designed for, is to analyze data from ChIP-seq experiment with a transcription factor of interest [94].

Table 5.4 shows top ten enriched molecular function term in common peak set from replicates and peak calling algorithms. The one with best score (lowest p-value) is related to Rab GTPase activity. Rab GTPases are part of Ras superfamily. The top 3 enriched molecular function are all involved in membrane traffic, such as vesicle formation, interacting with cytoskeleton for vesicle transfer.

Table 5.4. Enriched Molecular Function related terms in common peaks dataset.

Enricher Molecular Function terms	**p-value (-log10)**
Rab GTPase activator activity	12.87
beta-tubulin binding	12.48
actin filament binding	12.19
ribonuclease activity	11.63
Rab GTPase binding	10.81
IgG binding	10.69
DNA helicase activity	9.98
uridylyltransferase activity	9.68
proteine serine/thereonine phosphatase activity	9.44
NF-κB inducing kinase activity	9.39

GREAT searches for enriched pathway related term partnering with Panther Pathway analysis and Molecular signature database (MSigDB). As it is shown in Table 5.5 and 5.6, in both databases Notch signalling pathway and PI3 Kinase pathway have

high scores and low p-values. PI3 kinase is one of the main deregulated pathways in melanoma (Section 1.2.2). According to Table 5.5 and 5.6 most of top enriched pathways are related to hallmarks of cancer.

Table 5.5. Top enriched pathway terms in Panther Pathway Analysis.

Panther (Protein ANalysis THrough Evolutionary Relationships) Pathway	
Pathway	**p-value (-log10)**
Notch Signaling Pathway	25.65
PI3 Kinase Pathway	13.18
P53 Pathway Feedback Loops2	12.63
VEGF signaling pathway	11.61
Insulin/IGF pathway	9.82

Table 5.6. Top enriched pathway terms in Molecular Signature Database.

MSigDB (Molecular Signatures DataBase) Pathway	
Pathway	**P value (-log10)**
Tyrosinase Receptor Kinase A signaling pathway related genes	28.39
Notch Signaling related genes	14.32
WNT signaling pathway related genes	12.34
PI3K/AKT Signaling related genes	11.96
Insulin Receptor pathway related genes	11.28

Another database which GREAT uses, is Gene Ontology analysis and related terms. Enriched biological processes in genes associated with common peaks dataset, is summarized in the Table 5.7 . There are a number of enriched biological process term related to cellular trafficking. Also terms related to ER and cellular stress are enriched in our common peak set. Furthermore in order to define IRF4- directly regulated genes, we study the ChIP-seq and RNA-seq data together.

Table 5.7. Top enriched GO biological process terms related to common peaks dataset.

GO Biological Process	**p-value (-log10)**
Golgi vesicle transport	27.75
Response to endoplasmic reticulum stress	26.14
Response to unfolded protein	18.8
Vesicle docking involved in exocytosis	18.07
Notch Signaling Pathway	16.75
DNA integrity checkpoint	15.64
Positive regulation of RabGTPase activity	12.87

5.7. Integration of ChIP-seq and RNA-seq data

We can identify IRF4 directly regulated targets with ChIP-seq, however only with the help of RNA-seq data, we can comment about IRF4 has whether activating or repressing target genes or even have any effect on their transcription to begin with. Assigned genes from the ChIP-seq dataset were compared with available RNA-seq significant gene list [5]. As Figure 5.13 shows, there are 2143 common genes between two dataset. Additionally, to see if IRF4 plays more of an activator role or repressor role, we examined intersection of IRF4 activated genes verses ChIP-seq annotated genes. As it is shown in fig 5.14. These two dataset have 1105 genes in common. Furthermore, comparison of IRF4 repressed genes and ChIP-seq annotated genes showed 1038 common genes.

In order to elaborate more on the properties and roles of shared genes between two dataset, the mutual genes between ChIP-seq data and IRF4 activated gene list and also mutual genes between IRF4 repressed genes and ChIP-seq dataset, we performed GO analysis and visualization with GOrilla. [95]. The most enriched GO term for molecular function belongs to membrane and cellular trafficking and cell cycle regulations. Summarized Figures can be seen in Appendix I.

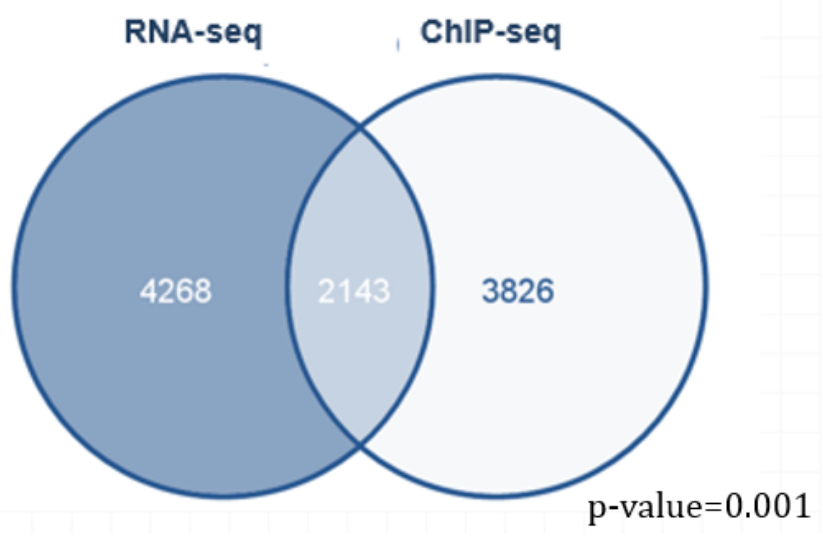


Figure 5.13. Venn diagram of common common genes between RNA-seq and ChIP-seq data.

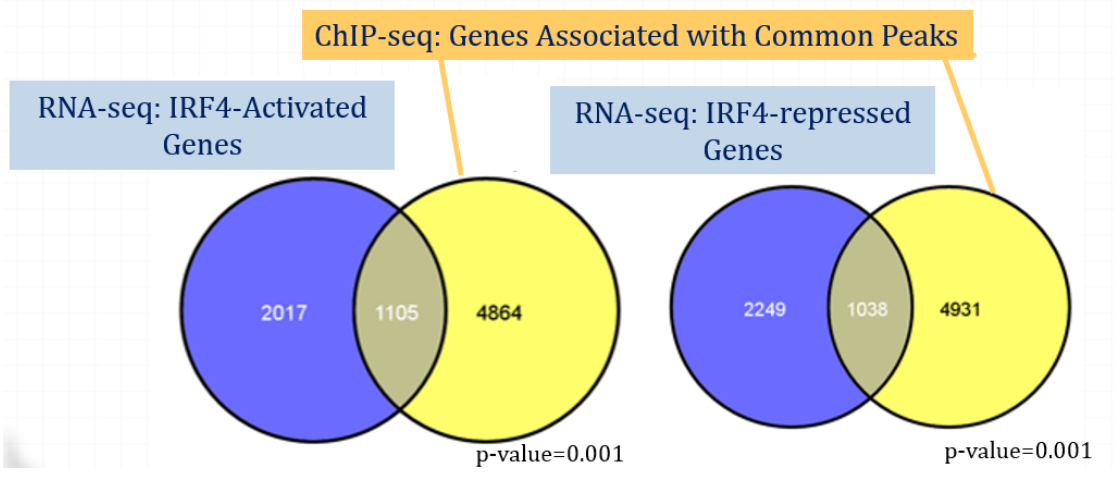


Figure 5.14. Venn diagram of common common genes between RNA-seq and ChIP-seq data.The left venn is just for IRF4 activated genes and the right venn diagram shows common genes which IRF4 represses them

5.8. ChIP-qPCR Verification of Selected Common Peaks from ChIP-seq Analysis

Integrating RNA-seq and ChIP-seq data, we have chosen some sample peaks region to be validated by qPCR. The selected peaks for qPCR were based on the quality of the peak which means these peaks are highly enriched compared to the background, present in common peaks and their associated gene is upregulated or downregulated according RNA-seq dataset. We have found a putative IRF4 binding sites on MITF, NRAS and TYR loci. In MITF locus, we observed several peaks however, the peak in Figure 5.15 was the one with highest enrichment; furthermore, considering RNA-seq results in Figure 5.18,the data suggests that MITF is directly regulated by IRF4.

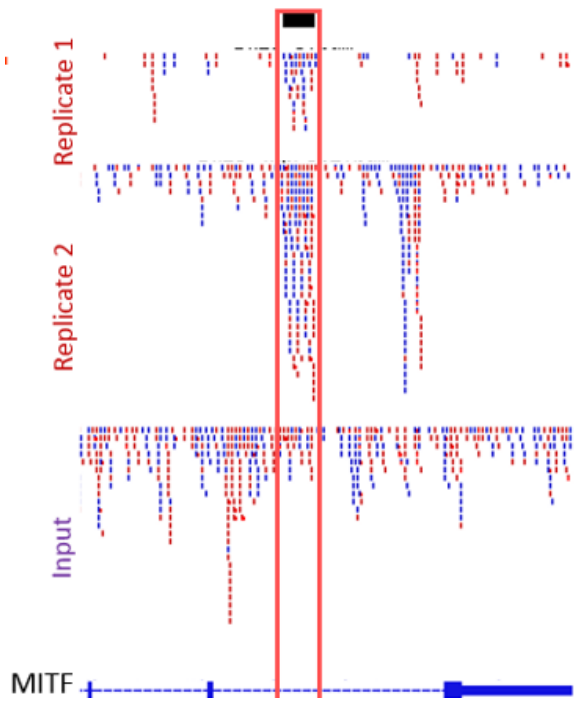


Figure 5.15. Putative IRF4 binding site in MITF gene. The red rectangle shows the possible IRF4 binding region.

According to both ChIP-seq and RNA-seq data as the Figures 5.16 and 5.18 demonstrates that IRF4 is directly regulating NRAS. Knocking-down IRF4 will also cause decrease in NRAS expression.

Figure 5.16. Putative IRF4 binding site in NRAS gene. The red rectangle shows the possible IRF4 binding region.

We also examined the peaks which were close to TSS and also were differentially expressed in RNA-seq data upon IRF4 knockdown. The chosen peak belongs to RAB3GAP1, which is membrane protein and part of Ras protein family. There are two putative IRF4 binding region for this gene. One of them is about 1kb upstream of TSS and the other peak 3kb downstream of the TSS in the first intronic region. We chose the one which is 1kb upstream of TSS for ChIP-qPCR verification (Figure 5.17). Furthermore, in RNA-seq data, we can observe a mild increase in RAB3GAP1 expression upon IRF4 knockdown.

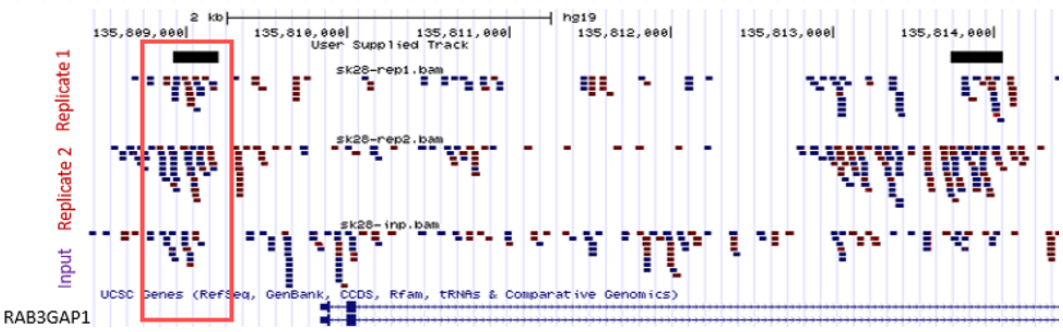


Figure 5.17. Putative IRF4 binding sites in RAB3GAP1 gene. The red rectangle shows one of the possible IRF4 binding regions.

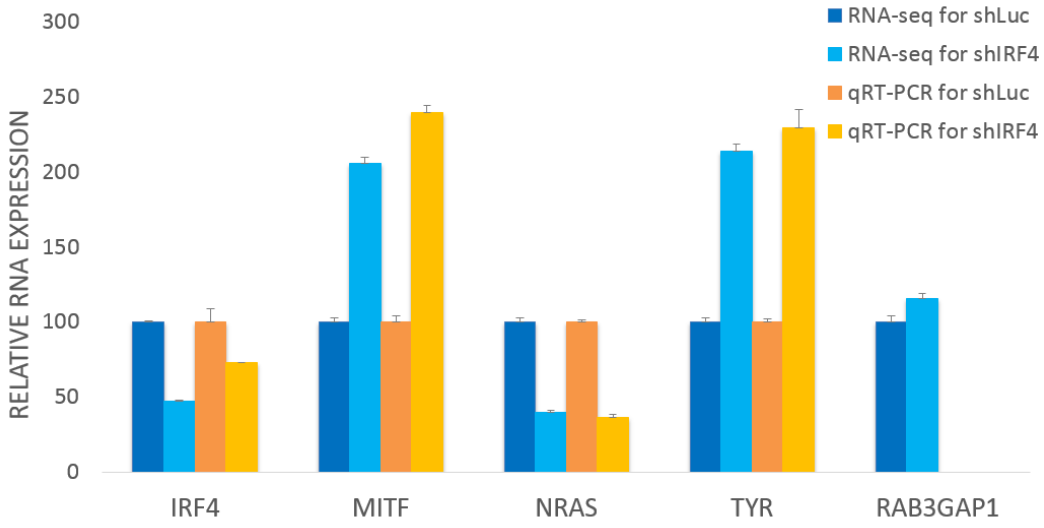


Figure 5.18. Expression of MITF, TYR, NRAS and RAB3GAP1 upon IRF4 knockdown. Except RAB3GAP1 gene, which does not have RT-qPCR data, all other three genes are presented with RNA-seq and RT-qPCR results. (Adopted from [5])

6. DISCUSSION

Transcriptional regulation by IRF4 has been extensively studied in immune cells and immune cell cancers, particularly lymphocytes and different types of lymphoma. IRF4 is essential for the survival and development of these cells. As previous studies have shown ABC-DLBCL and MM cells have a non-oncogenic addiction to IRF4 expression which makes IRF4 expression vital for survival and growth cancer cells [43,44]. Recent studies from our lab and elsewhere demonstrated that IRF4 expression is also present in melanocytes and melanoma cells [99]. Additionally, several genome-wide association studies linking a particular SNP in IRF4 locus is linked to predisposition to melanoma [67, 68]. Initial studies of IRF4 in melanoma cell lines, in our group, revealed possible non-oncogenic addiction of these cells to IRF4 such that downregulation of IRF4 expression in IRF4-positive melanoma cell lines led to loss of viability, and this may be because of a cell cycle arrest at G2/M checkpoint [99]. Studies from our lab and a recent article [4] all point out to IRF4 as a critical transcriptional regulator in melanocyte and melanomas. So one of the steps of elucidating functional roles of IRF4 in melanoma is to find its downstream transcriptional target genes. We set out to find the downstream transcriptional targets by two methods: ChIPseq and RNAseq. Yilmaz et al from our group, performed RNAseq and its analysis, and I performed ChIPseq and its analysis.

As the first step, we established and optimized different steps of chromatin immunoprecipitation protocol. First, we optimized sonication condition in order to achieve 200-700bp range in shearing chromatin of SkMel-28 lysates. The next steps of optimization involved balancing the concentration of salt and detergent in washing buffers in order to avoid unspecific binding, and last optimization step was designing and choosing the right primers for qPCR, in order to get amplicon which is unique to the putative IRF4 binding regions that also show good amplification in qPCR. Finally, in order to avoid low concentration of ChIP samples in library preparation, we pulled down four ChIP-DNA samples together as our starting amount at the beginning of library preparation.

Upon completion of ChIP-seq experiments, we started setting up our bioinformatics pipeline by quantifying per base read quality which were in satisfactory range for all samples. As a result, we only trimmed 5 base pair in total from the both ends of all sequenced reads.

In the mapping step , we used uniquely mapped reads for the downstream of analysis pipeline. We selected the uniquely mapped reads in order to avoid the uncertainty about the source of individual sequencing reads. Mapped reads in one of the replicates were almost half of the other replicate.

Following aligning the reads to genome, we checked the quality of ChIP-seq experiments with NGS-QC [87] to see if the we can observe the expected selective enrichment genome-wide or not. Our ChIP-seq replicates passed the analysis successfully. However, our input sample also showed some selective enrichment genome-wide. Therefore we decided to perform peak calling with three different computational algorithms: MACS, HOMER and BayesPeak. Furthermore, for more detailed analysis of the quality of ChIP-seq samples it would help to perform quality controls with other quality control computational algorithms designed for ChIP-seq such as ChIPQC of the Bioconductor.

In peak calling step, as expected, the replicate with lower read number also had lower peaks compared to the other replicate. Moreover, shared peaks between two replicates were about 50% in MACS, 22% in Homer and 28% in BayesPeak. To be certain that this difference is due to low number of reads, we tested this by randomly sampling 50-60% of the genome three times and then checked the common peaks between the two samples to see if the percentage of common peaks changed between the two samples. By lowering the number of reads, we achieved 50-60% of common peaks between two samples in the analysis using Homer suggesting that there are putative peaks in the sample with lower number of reads which were not recognized because their enrichment score was lower and possibly their false discovery rate were higher than 0.1.

In order to understand how many more reads are needed in order to find all IRF4 binding regions of the genome, we randomly sampled from the replicate with higher number of reads, starting with just 5% of all the reads and gradually increasing until 90% of the reads. This random sampling was repeated 10 times per each fraction of the reads. Then peak calling was performed with MACS. At the end, graph was plotted based on the acquired peak numbers. According to the visual inspection of the graph, number of reads is close to saturation point and sequencing more reads is not expected to result in too many more peaks to be defined and therefore would not be cost-efficient. On the other hand, a detailed mathematical analysis can be expected to give a more quantitative assessment of this point.

Furthermore, as part of downstream analysis we analyzed the sequences of our selected peaks with MEME-ChIP in order to find enriched motifs. As it was shown in Figure 5.12, centrally enriched motif is an AT rich motif associated with HOX family and NFAT. IRF motif is enriched 20-40bp away from the center of the peaks grouped with TFAP2C, NF-κB, MEF2A and MEF2B together. A possible explanation for why IRF motif is not centrally enriched in this dataset may arise from the method which we find common peaks. In BEDtools, in order to find common peaks, it searches for intersection of the peaks from the two user-given inputs and cuts out the shared regions. Due to our pipeline, we perform this finding common peaks step 3 times: 1)Intersection of the replicates. 2)Intersection of of two peak-caller results and 3) Intersection of output of number 2 with a third peak caller results. Considering the difference between peak callers and their peak finding algorithms, because MACS calculates peaks based on the read numbers and their enrichment compared to input, and on the other hand BayesPeak more relies on the formation and structure of peaks compared to input, and HOMER considers both read number and their possible peak structure, so it can be expected to see common peaks not to be the center of the peaks generated by each algorithm. Therefore, we performed motif analysis on top 25% peaks found by each algorithm. As it is demonstrated in Appendices H, I and J, MACS and HOMER show enriched IRF motif, and BayesPeak only has part of IRF motif as enriched motif (5′-GAA(A)-3′). Moreover, all the three algorithms contain AT-rich motifs. Another possible reason why IRF motif is not centrally enriched and got lower score that the

other motifs found in common peaks can be the way MEME-ChIP calculates and finds motifs. MEME erases the site that match previously found motif. Therefore, combining the BEDtools results and algorithm of MEME, we suggest that it can be expected not to see IRF motif enriched exactly at the center of the peaks.

For annotation and analysis of Gene Ontology terms and pathways in ChIP-seq data, we used GREAT because DAVID and GO::TermFinder, the two popular annotation tools, is not suitable for analysis of ChIP-seq data. These annotation enrichment tools are gene-based. This is not accurate because gene-based tests do not account for biases during assigning genomic regions to genes. A random genomic region is more likely to be assigned to a gene in a gene desert simply because deserts provide large regions where that gene is the nearest one. GREAT models this situation. Hence, more accurate enrichments for a set of genomic regions is calculated. GREAT also recruits and uses numerous ontologies providing a range of annotations such as protein domains and pathways [25, 94].

As a verification of ChIP-seq, we chose four peaks for validation with ChIP-qPCR. Three of these peaks are associated with MITF, NRAS and TYR, which play key roles in growth and survival of melanoma cells. The last chosen peak is based on two criteria, first unlike other peaks it is closer to transcription start site (maximum 1500bp upstream or downstream of TSS) and second criteria is that it scored high and have shown one of the best p-values and enrichments. This peak is aligned to about 1kb upstream of TSS of RAB3GAP1, a subunit of Rab GTPase which is member of Ras protein superfamily. RabGTPase is responsible for membrane trafficking and vesicle formation and transportation.

After we determined peaks which is common with both replicates and all three algorithms, we performed Gene Ontology analysis for these peaks. Later on, we checked for common genes between IRF4 ChIP-seq chosen peaks and IRF4 regulated significant gene list from previous RNA-seq results. They shared 2143 genes together, 1105 genes were in common with IRF4-activated gene list and 1038 genes with IRF4-repressed genes. In order to know more about these common genes, we used GOrilla to check for

gene ontology term of biological process, molecular function and cellular components [95]. We visualized the GOrilla data with REVIGO. As a result of the GOrilla analysis and REVIGO graphical output, enriched GO-BP terms of common genes of IRF4 ChIP-seq and IRF4-activated genes in SkMel-28 suggest an association to regulation of cyclin-dependent protein serine/thereonine kinase activity which suggests regulating cell cycle control and proliferation.

Gene ontology enrichment analysis with common genes of IRF4 ChIP-seq and IRF4-repressed gene lists of SkMel-28, we observe GO terms related to adenylate cyclases which are inhibitor of G-protein coupled receptor pathway, transition of metal-ion transport which are both related to membrane traffic system, and cellular transport. This is interesting because in the over-represented GO terms of IRF4-repressed genes in SkMel-28, we have also seen related enriched terms with vesicle mediated transport and melanin synthesis which suggests IRF4 may have an effect on melanosome formation and transport. Also enriched GO terms related to $\beta - tubulin$ and actin filament binding gives us more clues about role of IRF4 in regulating vesicle mediated transports across the cell.

As it is shown in Figure 5.13, 3826 genes which were associated between peaks did not stand out in RNA-seq results. Different reasons can give rise to this difference. One of these reasons can be because of how peaks are associated with genes. Unlike RNA-seq, in ChIP-seq most of the peaks associated genes are between 5-50kb upstream or downstream of TSS (Figure K.1). However these possible binding regions could be regulating genes far away on the same chromosome or other chromosomes. Unfortunately ChIP-seq does not give any information about long-range chromatin interactions of IRF4. Chromatin Conformataion capture techniques should be performed in order to find these interaction sites. Another reason of why have 3826 genes not common with RNA-seq may also be due to the fact that IRF4 has a lot of cofactors. Therefore, IRF4 alone may not be regulating these genes. Discovery of IRF4 co-factors in melanoma can help with characterization of these genes and in what conditions IRF4 regulates them.

ChIP-seq and RNA-seq results both demonstrate the regulatory role of IRF4 on MITF in melanoma. MITF is the master regulator in melanocyte development and it also has been shown to be involved in proliferation, survival, pigmentation, metastasis and invasion in melanoma cells [22]. We found a putative IRF4 binding site inside MITF in an intronic region. Furthermore, previous studies also give us clues about regulatory mechanism between MITF and IRF4. Hoek et al demonstrated that overexpression of MITF leads to upregulation of IRF4 in SkMel-28 cell line [3]. Also in a recent study, it was shown that MITF binds to intron 4 of IRF4 and in the case of downregulation of MITF, there is also decrease in IRF4 mRNA and protein levels [4]. Also the same study indicated that upon reduction in IRF4 levels in 501mel cell line, there is no change in MITF expression [4]. However MITF was one of the common genes between ChIP-seq annotated genes and RNA-seq IRF4-activated gene list, which suggests MITF-IRF4 relationship can vary among the cell lines. Also it might be a clue to other epigenetic factors and transcription factors which interacts with IRF4 during regulation of MITF and depending on the state of the cells, IRF4 may have roles in both repress and activation of melanoma. Another interesting common gene between ChIP-seq data and IRF4-activated gene set is NRAS gene suggesting that NRAS is likely regulated directly by IRF4. NRAS is one of the key proteins in melanoma. It is a mediator of MAPK/ERK and PI3K/AKT pathways which are over-activated in most of the melanoma cells. Also according to analysis done using GREAT, PI3K/AKT signaling pathway is one of the over-represented term in both Panther and MSigDB pathway analysis. This findings is also consistent with Yılmaz observations. In our lab, studies have shown that p-AKT levels (p-AKT (Ser473) and p-AKT (Thr308), two sites phosphorylated by mTORC2 and PDK1, respectively) in SkMel-28 are reduced upon IRF4 knock-down and increased when IRF4 is overexpressed. All these evidences suggests that IRF4 may effect AKT activation at least partly through regulating NRAS. In brief, after carrying out a comprehensive analysis of ChIP-seq experiments and its integration with RNA-seq data, the GO enrichment analysis of the mutual genes, IRF4 appears to play a key role in regulating cell cycle. Also, it seems that IRF4 play a major role in regulating cellular and membrane transport pathways such as melanosome transportation. Furthermore, there is still a lot of space to wander about

IRF4 roles in growth and survival of melanoma cell lines. For instance, it would be great to perform ChIP-seq in other IRF4 expressing melanoma cell lines. Considering the RNA-seq results and the differences between expression profiles of SkMel-28 and SkMel-5 upon IRF4 knockdown, it will be interesting to investigate if localization of IRF4 and its target genes changes among melanoma cell lines.

APPENDIX A: OPTIMIZATION OF CHIP

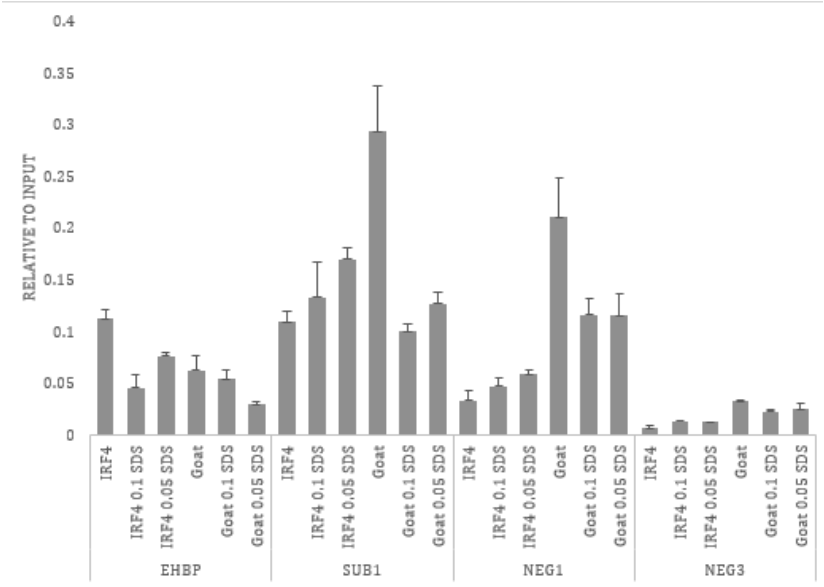


Figure A.1. Anti-IRF4 ChIP-qPCR Optimization of Immune-Complex washing buffers. IRF4 or Goat IgG samples were washed with buffers containing 0.1% and 0.05% SDS, respectively.

APPENDIX B: ANTI-IRF4 CHIP-qPCR ON TYROSINASE PROMOTER IN SKMEL-5 CELL LINE

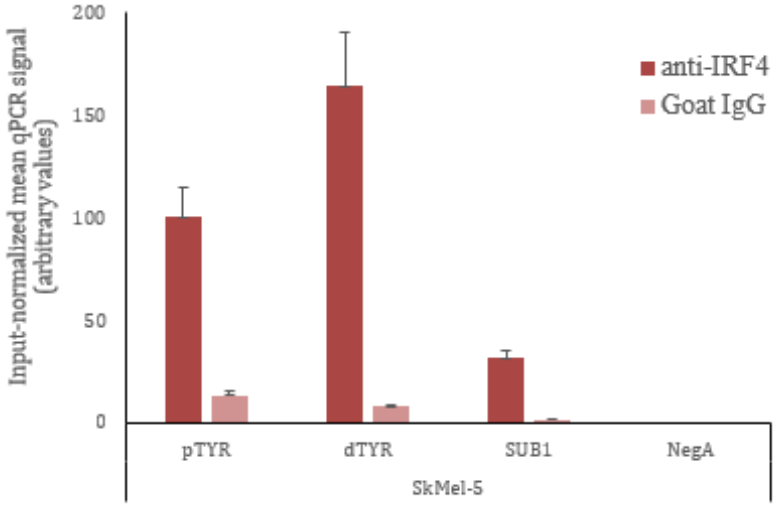


Figure B.1. Quantitative anti-IRF4 ChIP for TYR promoter in SkMel-5 cell line. The TYR primers are indicated as pTYR for proximal promoter region and dTYR for distal promoter region. SUB1 is the positive control primer pair and NegA is the negative control primer pair.

APPENDIX C: AGILENT BIOANALYZER PROFILE OF THE SAMPLES

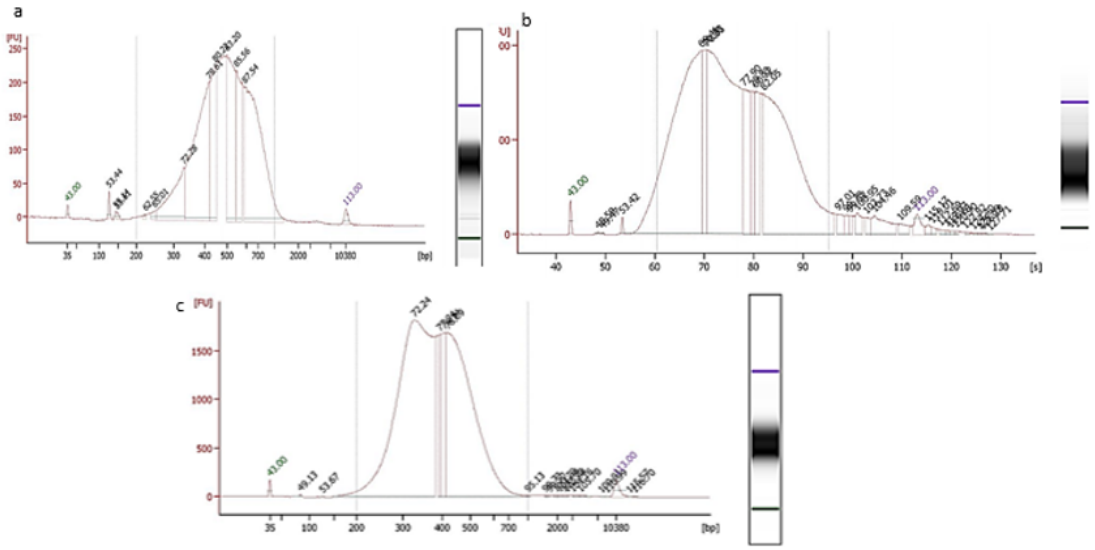


Figure C.1. Agilent Bioanalyzer produced graphs for the samples. This graphs shows average size of fragment after library preparation.

APPENDIX D: PER BASE QUALITY GRAPHS OF SKMEL-28 SAMPLES

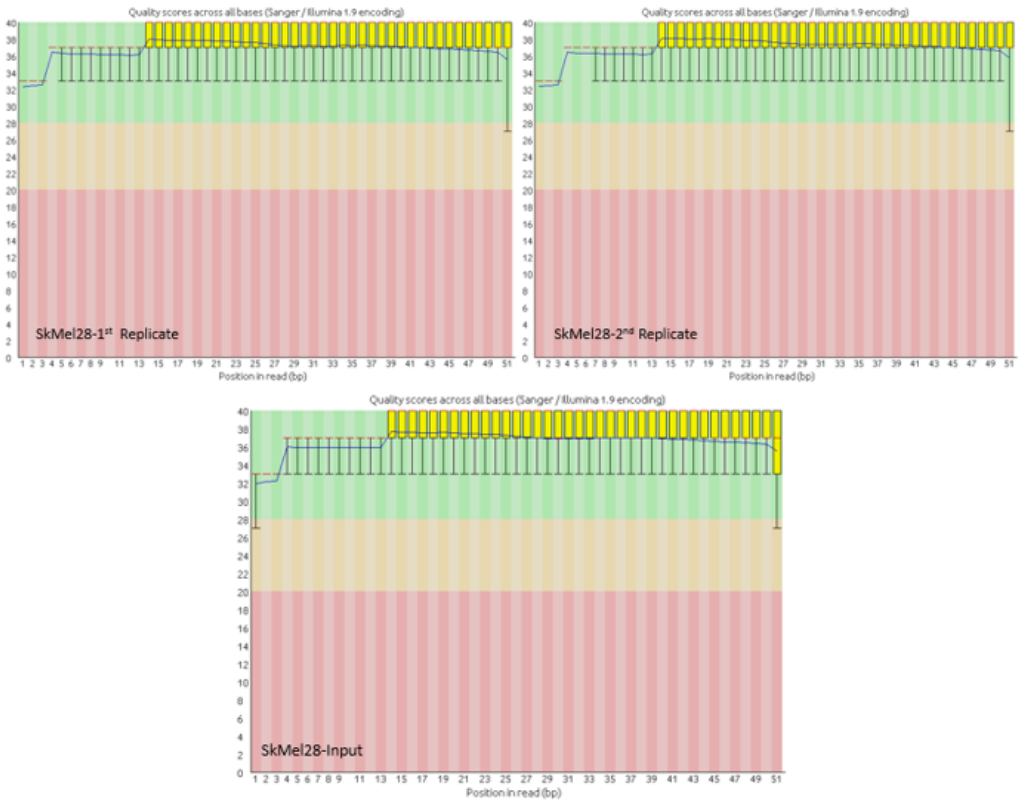


Figure D.1. Per base quality graphs for SkMel-28 replicates and input. x-axis shows base position and y-axis indicates a Phred quality score.

APPENDIX E: VISUALIZATION OF SKMEL-28 MAPPED READS

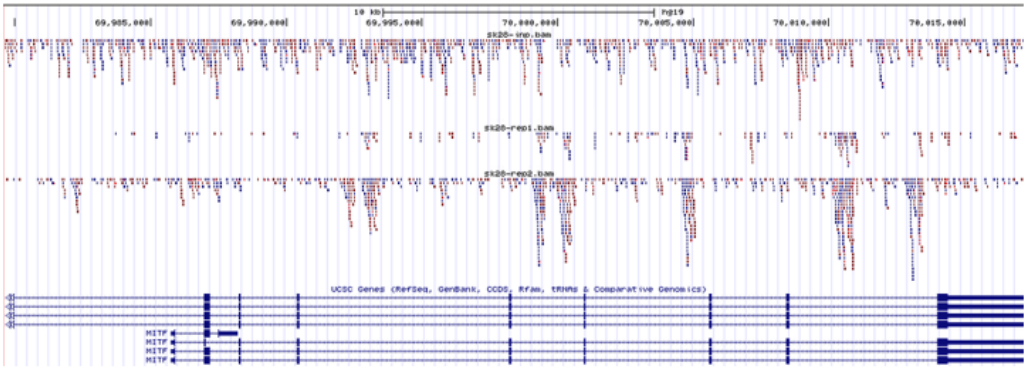


Figure E.1. UCSC Genome Browser visualization at MITF locus with reads aligned in stacked mode in ChIP samples and more dispersed in input. First row shows input of SkMel-28. Next two rows indicate reads in ChIP-seq replicates. Visualizations were performed with mapped reads without read number normalization between samples.

APPENDIX F: NGS-QC QUALITY CONTROLS OF SKMEL-28 SAMPLES

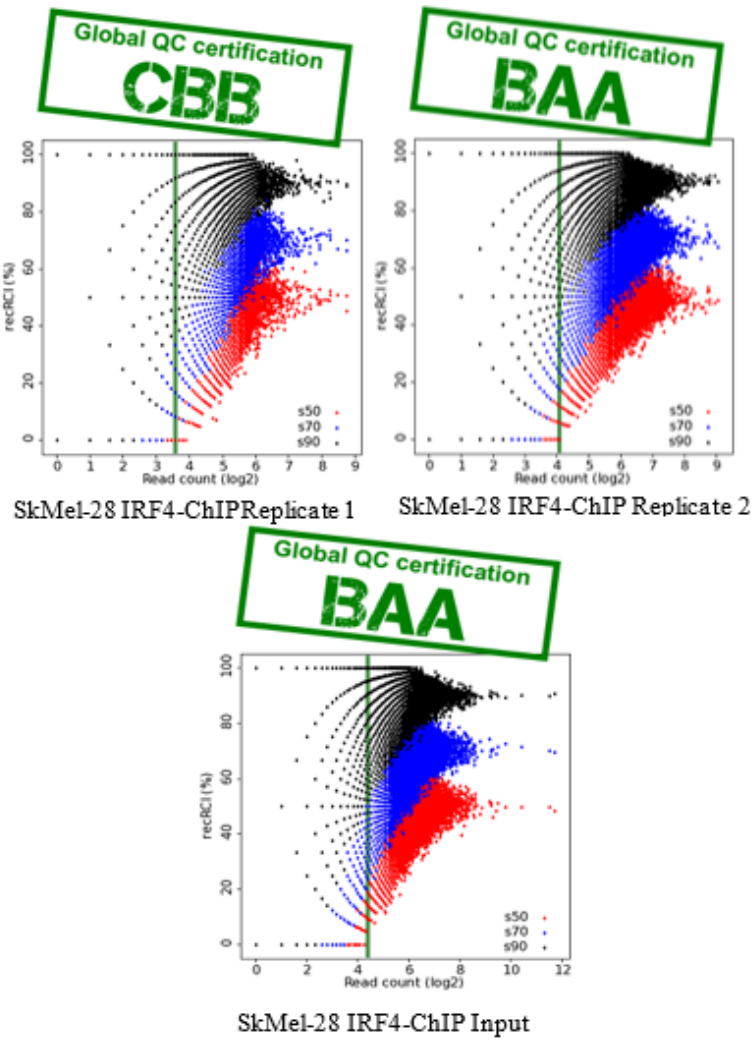


Figure F.1. NGS-QC Quality reports of the SkMel-28 samples. Effect of random sampling on the profile. This figure demonstrates the effect of the random sampling subsets (90%: black; 70%: blue; 50%:red) on the recovered read count Intensity (recRCI) per bin. The dark-green vertical line denotes the background threshold.

APPENDIX G: MACS: MODEL PEAKS

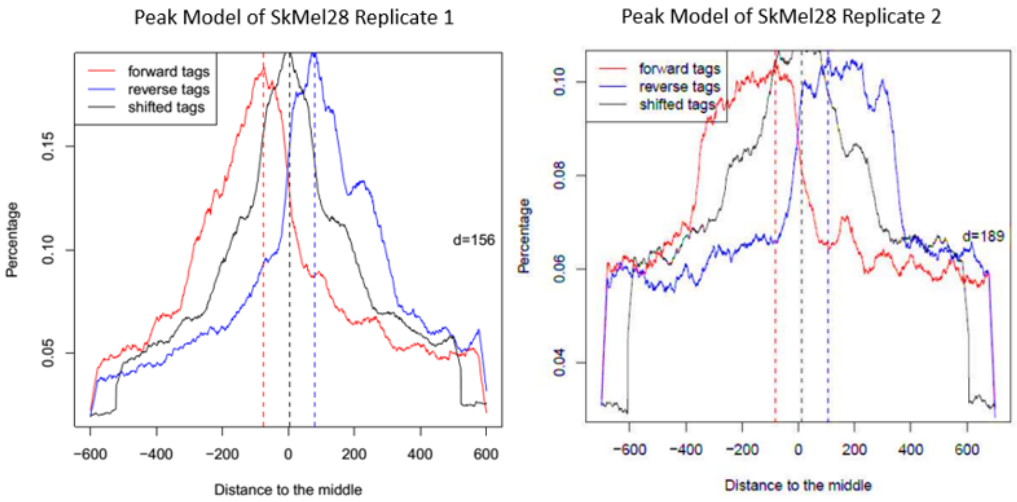


Figure G.1. Comparison of Peak model for replicate 1 vs replicate 2 of MACS analysis.

APPENDIX H: MOTIF ANALYSIS FOR MACS PEAKS

MOTIF	Known or similar motifs	E-Value
	TFAP2C/TFAP2B HINFP1	2.0e-1083
	-	6.5e-3423
	-	1.3e-2293
	NFAT5	3.1e-2231
	ASCL2	3.4e-1854
	ZNF524	1.4e-1843
	IRF7	4.5e-879
	RARA/THRA/RORA	6.8e-846

Figure H.1. Enriched motifs in Peaks found by MACS. E-value is the p-value multiplied by the number of motifs in the input database.

APPENDIX I: MOTIF ANALYSIS FOR HOMER PEAKS

MOTIF	Known or similar motifs	E-Value
	MEF2A	2.4e-5590
	FLI1	3.4e-4461
	MEF2D	2.5e-1583
	POU3F3	2.2e-941
	MEF2A-DBD	1.5e-684
	HIC1-DBD	3.6e-359
	TP53-DBD	1.0e-308
	ZNF524	1.2e-281
	YY2	1.7e-250
	IRF5	1.9e-069

Figure I.1. Enriched motifs in Peaks found by HOMER. E-value is the p-value multiplied by the number of motifs in the input database.

APPENDIX J: MOTIF ANALYSIS FOR BAYESPEAK PEAKS

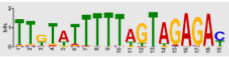
MOTIF	Known or similar motifs	E-Value
	-	2.7e-1166
	-	2.0e-3214
	MEF2D	4.5e-1414
	-	8.6e-2503
	NFAT5-DBD	8.0e-1942
	PAX9-DBD	5.0e-1399
	-	2.3e-682
	ZNF524	1.3e-288
	-	1.0e-231
	TFAP2B/TFAP2C	4.4e-140

Figure J.1. Enriched motifs in Peaks found by BayesPeak. E-value is the p-value multiplied by the number of motifs in the input database.

APPENDIX K: DISTANCE OF PEAKS FROM TRANSCRIPTION START SITES

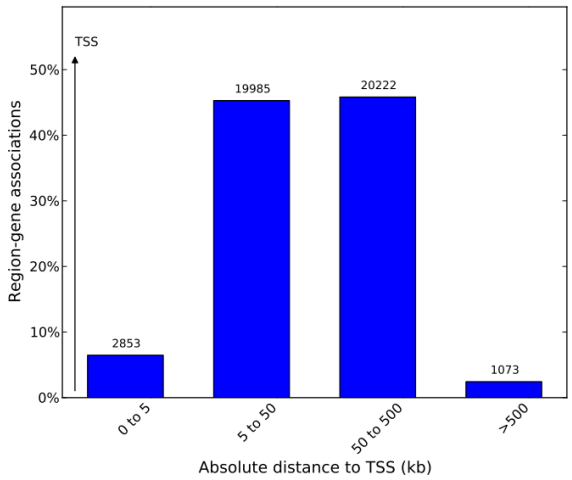


Figure K.1. Absolute distance of peaks from TSS.

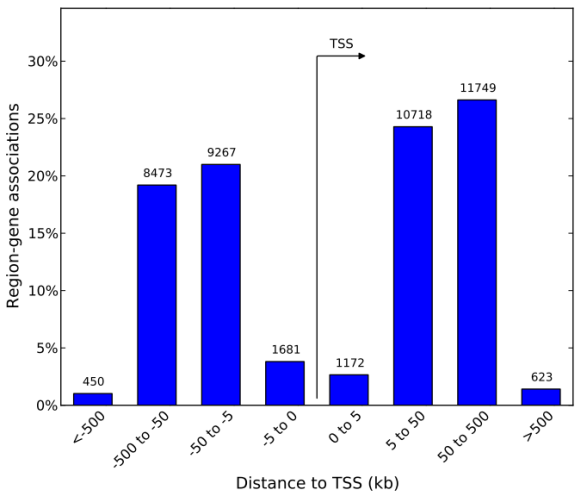


Figure K.2. Distance of peaks from TSS, both downstream and upstream of TSS.

REFERENCES

1. Hanahan, D. and R. a. Weinberg, "Hallmarks of Cancer: the Next Generation.", *Cell*, Vol. 144, No. 5, pp. 646–674, 2011.

2. Vultur, A. and M. Herlyn, "SnapShot: Melanoma.", *Cancer Cell*, Vol. 23, No. 5, p. 706, 2013.

3. Hoek, K. S., N. C. Schlegel, O. M. Eichhoff, D. S. Widmer, C. Praetorius, S. O. Einarsson, S. Valgeirsdottir, K. Bergsteinsdottir, A. Schepsky, R. Dummer and E. Steingrimsson, "Novel MITF Targets Identified Using a Two-step DNA Microarray Strategy.", *Pigment Cell & Melanoma Research*, Vol. 21, No. 6, pp. 665–676, 2008.

4. Praetorius, C., C. Grill, S. N. Stacey, A. M. Metcalf, D. U. Gorkin, K. C. Robinson, E. Van Otterloo, R. S. Q. Kim, K. Bergsteinsdottir, M. H. Ogmundsdottir, E. Magnusdottir, P. J. Mishra, S. R. Davis, T. Guo, M. R. Zaidi, A. S. Helgason, M. I. Sigurdsson, P. S. Meltzer, G. Merlino, V. Petit, L. Larue, S. K. Loftus, D. R. Adams, U. Sobhiafshar, N. C. T. Emre, W. J. Pavan, R. Cornell, A. G. Smith, A. S. McCallion, D. E. Fisher, K. Stefansson, R. a. Sturm and E. Steingrimsson, "A Polymorphism in IRF4 Affects Human Pigmentation Through a Tyrosinase-dependent MITF/TFAP2A Pathway.", *Cell*, Vol. 155, No. 5, pp. 1022–1033, 2013.

5. Erdem, Y., *Characterization of Interferon Regulatory Factor 4 (IRF4) Target Genes in Melanoma Cell Lines*, Master's Thesis, Bogazici University, 2014.

6. Hanahan, D., R. A. Weinberg and S. Francisco, "The Hallmarks of Cancer", *Cell*, Vol. 100, pp. 57–70, 2000.

7. Lemmon, M. a. and J. Schlessinger, "Cell Signaling by Receptor Tyrosine Kinases.", *Cell*, Vol. 141, No. 7, pp. 1117–1134, 2010.

8. Jiang, B.-H. and L.-Z. Liu, "PI3K/PTEN Signaling in Angiogenesis and Tumorigenesis.", *Advances in Cancer Research*, Vol. 102, No. 09, pp. 19–65, 2009.

9. Ghebranious, N. and L. A. Donehower, "Mouse mModels in Tumor Suppression", *Oncogene*, , No. 17, pp. 3385–3400, 1998.

10. Adams, J. M. and S. Cory, "Bcl-2-Regulated Apoptosis: Mechanism and Therapeutic Potential.", *Current Opinion in Immunology*, Vol. 19, No. 5, pp. 488–496, 2007.

11. Berx, G. and F. van Roy, "Involvement of Members of the Cadherin Superfamily in Cancer.", *Cold Spring Harbor Perspectives in Biology*, Vol. 1, No. 6, pp. 1–27, 2009.

12. Micalizzi, D. S., S. M. Farabaugh and H. L. Ford, "Epithelial-mesenchymal Transition in Cancer: Parallels Between Normal Development and Tumor Progression.", *Journal of Mammary Gland Biology and Neoplasia*, Vol. 15, No. 2, pp. 117–1134, 2010.

13. Schadendorf, D. and A. Hauschild, "Melanoma in 2013: Melanoma– The Run of Success Continues.", *Nature Reviews. Clinical Oncology*, Vol. 11, No. 2, pp. 75–76, 2014.

14. Davies, M. a. and Y. Samuels, "Analysis of the Genome to Personalize Therapy for Melanoma.", *Oncogene*, Vol. 29, No. 41, pp. 5545–5555, 2010.

15. Ekedahl, H., H. Cirenajwis, K. Harbst, a. Carneiro, K. Nielsen, H. Olsson, L. Lundgren, C. Ingvar and G. Jönsson, "The Clinical Significance of BRAF and NRAS Mutations in a Clinic-based Metastatic Melanoma Cohort.", *The British Journal of Dermatology*, Vol. 169, No. 5, pp. 1049–1055, 2013.

16. Gray-Schopfer, V. C., S. C. Cheong, H. Chong, J. Chow, T. Moss, Z. a. Abdel-Malek, R. Marais, D. Wynford-Thomas and D. C. Bennett, "Cellular Senescence

in Naevi and Immortalisation in Melanoma: a Role for p16?", *British Journal of Cancer*, Vol. 95, No. 4, pp. 496–505, 2006.

17. Hodis, E., I. R. Watson, G. V. Kryukov, S. T. Arold, M. Imielinski, J.-P. Theurillat, E. Nickerson, D. Auclair, L. Li, C. Place, D. Dicara, A. H. Ramos, M. S. Lawrence, K. Cibulskis, A. Sivachenko, D. Voet, G. Saksena, N. Stransky, R. C. Onofrio, W. Winckler, K. Ardlie, N. Wagle, J. Wargo, K. Chong, D. L. Morton, K. Stemke-Hale, G. Chen, M. Noble, M. Meyerson, J. E. Ladbury, M. a. Davies, J. E. Gershenwald, S. N. Wagner, D. S. B. Hoon, D. Schadendorf, E. S. Lander, S. B. Gabriel, G. Getz, L. a. Garraway and L. Chin, "A Landscape of Driver Mutations in Melanoma.", *Cell*, Vol. 150, No. 2, pp. 251–263, 2012.

18. Pollock, P. M., U. L. Harper, K. S. Hansen, L. M. Yudt, M. Stark, C. M. Robbins, T. Y. Moses, G. Hostetter, U. Wagner, J. Kakareka, G. Salem, T. Pohida, P. Heenan, P. Duray, O. Kallioniemi, N. K. Hayward, J. M. Trent and P. S. Meltzer, "High frequency of BRAF Mutations in Nevi.", *Nature Genetics*, Vol. 33, No. 1, pp. 19–20, 2003.

19. Bertolotto, C., "Metabolism Under the Spotlight in Senescence.", *Pigment Cell & Melanoma Research*, Vol. 27, No. 1, pp. 3–5, 2014.

20. Ackermann, J., M. Frutschi, K. Kaloulis, T. Mckee, A. Trumpp and F. Beermann, "Metastasizing Melanoma Formation Caused by Expression of Activated N-Ras Q61K on an INK4a-Deficient Background", *Cancer Research*, Vol. 65, No. 10, pp. 4005–4012, 2005.

21. Cheli, Y., M. Ohanna, R. Ballotti and C. Bertolotto, "Fifteen-year Quest for Microphthalmia-Associated Transcription Factor Target Genes.", *Pigment Cell & Melanoma Research*, Vol. 23, No. 1, pp. 27–40, 2010.

22. Levy, C., M. Khaled and D. E. Fisher, "MITF: Master Regulator of Melanocyte Development and Melanoma Oncogene.", *Trends in Molecular Medicine*, Vol. 12, No. 9, pp. 406–414, 2006.

23. Garraway, L. a., H. R. Widlund, M. a. Rubin, G. Getz, A. J. Berger, S. Ramaswamy, R. Beroukhim, D. a. Milner, S. R. Granter, J. Du, C. Lee, S. N. Wagner, C. Li, T. R. Golub, D. L. Rimm, M. L. Meyerson, D. E. Fisher and W. R. Sellers, "Integrative Genomic Analyses Identify MITF as a Lineage Survival Oncogene Amplified in Malignant Melanoma.", *Nature*, Vol. 436, No. 7047, pp. 117–122, 2005.

24. Carreira, S., J. Goodall, L. Denat, M. Rodriguez, P. Nuciforo, K. S. Hoek, A. Testori, L. Larue and C. R. Goding, "Mitf Regulation of Dia1 Controls Melanoma Proliferation And Invasiveness", *Genes & Development*, Vol. 20, pp. 3426–3439, 2006.

25. Boyle, G. M., S. L. Woods, V. F. Bonazzi, M. S. Stark, E. Hacker, L. G. Aoude, K. Dutton-Regester, A. L. Cook, R. a. Sturm and N. K. Hayward, "Melanoma Cell Invasiveness is Regulated by miR-211 Suppression of the BRN2 Transcription Factor.", *Pigment Cell & Melanoma Research*, Vol. 24, No. 3, pp. 525–537, 2011.

26. Yamagata, T., J. Nishida, T. Tanaka, R. Sakai, K. Mitani, M. Yoshida, T. Taniguchi, Y. Yazaki and H. Hirai, "A Novel Interferon Regulatory Factor Family Transcription Factor , ICSAT / Pip / LSIRF , That Negatively Regulates the Activity of Interferon-Regulated Genes", *Molecular and Cellular Biology*, Vol. 16, No. 4, pp. 1283–1294, 1996.

27. Mittrucker, H.-w., T. Matsuyama, A. Grossman, T. M. Ku, J. Potter, A. Shahinian, A. Wakeham, B. Patterson, P. S. Ohashi and T. W. Mak, "Requirement for the Transcription Factor LSIRF / IRF4 for Mature B and T Lymphocyte Function", *Science*, Vol. 275, pp. 1–4, 1997.

28. Gupta, B. S., M. Jiang, A. Anthony and A. B. Pernis, "Lineage-specific Modulation of Interleukin 4 Signaling by Interferon Regulatory Factor 4", *Journal of Experimental Medicine*, Vol. 190, No. 12, pp. 1837–1848, 1999.

29. Shaffer, a. L., A. Rosenwald and L. M. Staudt, "Lymphoid Malignancies: The

Dark Side of B-cell Differentiation.", *Nature Reviews. Immunology*, Vol. 2, No. 12, pp. 920–932, 2002.

30. Lin, L., A. J. Gerth and S. L. Peng, "Active Inhibition of Plasma Cell Development in Resting B Cells by Microphthalmia-associated Transcription Factor.", *The Journal of Experimental Medicine*, Vol. 200, No. 1, pp. 115–122, 2004.

31. Brüstle, A., S. Heink, M. Huber, C. Rosenplänter, C. Stadelmann, P. Yu, E. Arpaia, T. W. Mak, T. Kamradt and M. Lohoff, "The Development of Inflammatory T(H)-17 Cells Requires Interferon-Regulatory Factor 4. ", *Nature Immunology*, Vol. 8, No. 9, pp. 958–966, 2007.

32. Honma, K., D. Kimura, N. Tominaga, M. Miyakoda, T. Matsuyama and K. Yui, "Interferon Regulatory Factor 4 Differentially Regulates the Production of Th2 Cytokines in Naïve vs. Effector / Memory CD4 T Cells", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 105, No. 41, pp. 15890–15895, 2008.

33. Brass, a. L., a. Q. Zhu and H. Singh, "Assembly Requirements of PU.1-Pip (IRF-4) Activator Complexes: Inhibiting Function in Vivo Using Fused Dimers.", *The EMBO Journal*, Vol. 18, No. 4, pp. 977–991, 1999.

34. Eisenbeis, C. F., H. Singh and U. Storb, "Pip, a Novel IRF Family Member, is a Lymphoid-specific, PU.1-dependent Transcriptional Activator.", *Genes & Development*, Vol. 9, No. 11, pp. 1377–1387, 1995.

35. Pernis, A. B., "The Role of IRF-4 in B and T Cell Activation and Differentiation.", *Journal of Interferon & Cytokine Research : The Official Journal of the International Society for Interferon and Cytokine Research*, Vol. 22, No. 1, pp. 111–120, 2002.

36. Escalante, C. R., A. L. Brass, J. M. Pongubala, E. Shatova, L. Shen, H. Singh and A. K. Aggarwal, "Crystal Structure of PU.1/IRF-4/DNA Ternary Complex",

Molecular Cell, Vol. 10, No. 5, pp. 1097–1105, 2002.

37. Atchison, M. L., "Transcription Factor Pip Can Enhance DNA Binding by E47 , Leading to Transcriptional Synergy Involving Multiple Protein Domains", *Molecular and Cellular Biology*, Vol. 18, No. 8, pp. 4639–4650, 1998.

38. Rengarajan, J., "Interferon Regulatory Factor 4 (IRF4) Interacts with NFATc2 to Modulate Interleukin 4 Gene Expression", *Journal of Experimental Medicine*, Vol. 195, No. 8, pp. 1003–1012, 2002.

39. Li, P., R. Spolski, W. Liao, L. Wang, T. L. Murphy, K. M. Murphy and W. J. Leonard, "BATF-JUN is Critical for IRF4-mediated Transcription in T Cells.", *Nature*, Vol. 490, No. 7421, pp. 543–546, 2012.

40. Iida, S., P. H. Rao, M. Butler, P. Corradini, M. Boccadoro, B. Klein, R. S. Chaganti and R. Dalla-Favera, "Deregulation of MUM1/IRF4 by Chromosomal Translocation in Multiple Myeloma.", *Nature Genetics*, Vol. 17, No. 2, pp. 226–230, 1997.

41. Lenz, G., I. Nagel, R. Siebert, A. V. Roschke, W. Sanger, G. W. Wright, S. S. Dave, B. Tan, H. Zhao, A. Rosenwald, H. K. Muller-Hermelink, R. D. Gascoyne, E. Campo, E. S. Jaffe, E. B. Smeland, R. I. Fisher, W. M. Kuehl, W. C. Chan and L. M. Staudt, "Aberrant Immunoglobulin Class Switch Recombination and Switch Translocations in Activated B cell-like Diffuse Large B Cell Lymphoma.", *The Journal of Experimental Medicine*, Vol. 204, No. 3, pp. 633–643, 2007.

42. Lenz, G., G. W. Wright, N. C. T. Emre, H. Kohlhammer, S. S. Dave, R. E. Davis, S. Carty, L. T. Lam, a. L. Shaffer, W. Xiao, J. Powell, A. Rosenwald, G. Ott, H. K. Muller-Hermelink, R. D. Gascoyne, J. M. Connors, E. Campo, E. S. Jaffe, J. Delabie, E. B. Smeland, L. M. Rimsza, R. I. Fisher, D. D. Weisenburger, W. C. Chan and L. M. Staudt, "Molecular Subtypes of Diffuse Large B-cell Lymphoma Arise by Distinct Genetic Pathways.", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 105, No. 36, pp. 13520–13525, 2008.

43. Yang, Y., A. L. Shaffer, N. C. T. Emre, M. Ceribelli, M. Zhang, G. Wright, W. Xiao, J. Powell, J. Platig, H. Kohlhammer, R. M. Young, H. Zhao, Y. Yang, W. Xu, J. J. Buggy, S. Balasubramanian, L. a. Mathews, P. Shinn, R. Guha, M. Ferrer, C. Thomas, T. a. Waldmann and L. M. Staudt, "Exploiting Synthetic Lethality for the Therapy of ABC Diffuse Large B Cell Lymphoma.", *Cancer Cell*, Vol. 21, No. 6, pp. 723–737, 2012.

44. Shaffer, A. L., N. C. T. Emre, L. Lamy, V. N. Ngo, G. Wright, W. Xiao, J. Powell, S. Dave, X. Yu, H. Zhao, Y. Zeng, B. Chen, J. Epstein and L. M. Staudt, "IRF4 Addiction in Multiple Myeloma.", *Nature*, Vol. 454, No. 7201, pp. 226–231, 2008.

45. Shaffer, A. L., N. C. T. Emre, P. B. Romesser and L. M. Staudt, "IRF4: Immunity. Malignancy! Therapy?", *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, Vol. 15, No. 9, pp. 2954–2961, 2009.

46. Acquaviva, J., X. Chen and R. Ren, "IRF-4 Functions as a Tumor Suppressor in Early B-cell Development.", *Blood*, Vol. 112, No. 9, pp. 3798–806, 2008.

47. Pathak, S., S. Ma, L. Trinh, J. Eudy, K.-U. Wagner, S. S. Joshi and R. Lu, "IRF4 is a Suppressor of c-Myc Induced B Cell Leukemia.", *PloS One*, Vol. 6, No. 7, p. e22628, 2011.

48. Teng, Y., Y. Takahashi, M. Yamada, T. Kurosu, T. Koyama, O. Miura and T. Miki, "IRF4 Negatively Regulates Proliferation of Germinal Center B cell-derived Burkitt's Lymphoma Cell Lines and Induces Differentiation Toward Plasma Cells.", *European Journal of Cell Biology*, Vol. 86, No. 10, pp. 581–589, 2007.

49. Ma, S., S. Pathak, L. Trinh and R. Lu, "Interferon Regulatory Factors 4 and 8 Induce the Expression of Ikaros and Aiolos to Down-regulate Pre – B-cell Receptor and Promote Cell-cycle Withdrawal in Pre – B-cell Development", *Blood*, Vol. 111, No. 3, pp. 1396–1404, 2008.

50. Lehtonen, a., V. Veckman, T. Nikula, R. Lahesmaa, L. Kinnunen, S. Matikainen

and I. Julkunen, "Differential Expression of IFN Regulatory Factor 4 Gene in Human Monocyte-Derived Dendritic Cells and Macrophages", *The Journal of Immunology*, Vol. 175, No. 10, pp. 6570–6579, 2005.

51. Satoh, T., O. Takeuchi, A. Vandenbon, K. Yasuda, Y. Tanaka, Y. Kumagai, T. Miyake, K. Matsushita, T. Okazaki, T. Saitoh, K. Honma, T. Matsuyama, K. Yui, T. Tsujimura, D. M. Standley, K. Nakanishi, K. Nakai and S. Akira, "The Jmjd3-Irf4 Axis Regulates M2 Macrophage Polarization and Host Responses Against Helminth Infection.", *Nature Immunology*, Vol. 11, No. 10, pp. 936–944, 2010.

52. Tamura, T., P. Tailor, K. Yamaoka, H. J. Kong, H. Tsujimura, J. J. O'Shea, H. Singh and K. Ozato, "IFN Regulatory Factor-4 and -8 Govern Dendritic Cell Subset Development and Their Functional Diversity", *The Journal of Immunology*, Vol. 174, No. 5, pp. 2573–2581, 2005.

53. Persson, E. K., H. Uronen-Hansson, M. Semmrich, A. Rivollier, K. Hägerbrand, J. Marsal, S. Gudjonsson, U. Hå kansson, B. Reizis, K. Kotarsky and W. W. Agace, "IRF4 Transcription-factor-dependent CD103(+)CD11b(+) Dendritic Cells Drive Mucosal T Helper 17 Cell Differentiation.", *Immunity*, Vol. 38, No. 5, pp. 958–969, 2013.

54. Gao, Y., S. a. Nish, R. Jiang, L. Hou, P. Licona-Limón, J. S. Weinstein, H. Zhao and R. Medzhitov, "Control of T Helper 2 responses by Transcription Factor IRF4-dependent Dendritic Cells.", *Immunity*, Vol. 39, No. 4, pp. 722–732, 2013.

55. Williams, J. W., M. Y. Tjota, B. S. Clay, B. Vander Lugt, H. S. Bandukwala, C. L. Hrusch, D. C. Decker, K. M. Blaine, B. R. Fixsen, H. Singh, R. Sciammas and A. I. Sperling, "Transcription Factor IRF4 Drives Dendritic Cells to Promote Th2 Differentiation.", *Nature Communications*, Vol. 4, No. May, p. 2990, 2013.

56. Eguchi, J., Q.-W. Yan, D. E. Schones, M. Kamal, C.-H. Hsu, M. Q. Zhang, G. E. Crawford and E. D. Rosen, "Interferon Regulatory Factors are Transcriptional

Regulators of Adipogenesis.", *Cell Metabolism*, Vol. 7, No. 1, pp. 86–94, 2008.

57. Jiang, D.-S., Z.-Y. Bian, Y. Zhang, S.-M. Zhang, Y. Liu, R. Zhang, Y. Chen, Q. Yang, X.-D. Zhang, G.-C. Fan and H. Li, "Role of Interferon Regulatory Factor 4 in the Regulation of Pathological Cardiac Hypertrophy.", *Hypertension*, Vol. 61, No. 6, pp. 1193–1202, 2013.

58. Guo, S., Z.-Z. Li, D.-S. Jiang, Y. Y. Lu, Y. Liu, L. Gao, S.-M. Zhang, H. Lei, L.-H. Zhu, X.-D. Zhang, D.-P. Liu and H. Li, "IRF4 is a Novel Mediator for Neuronal Survival in Ischaemic Stroke.", *Cell Death and Differentiation*, Vol. 21, No. 6, pp. 888–903, 2014.

59. Grossman, A., H. W. Mittrücker, J. I. N. Nicholl, A. K. S. Suzuki, S. Chung, L. Antonio, S. Suggs, G. R. Sutherland, D. P. Siderovski and T. W. Mak, "Cloning of Human Lymphocyte-Specific Interferon Regulatory Factor (hLSIRF / hIRF4) and Mapping of the Gene to 6p23 – p25", *Genomics*, , No. 37, pp. 229–233, 1996.

60. Rhodes, D. R., J. Yu, K. Shanker, N. Deshpande and R. Varambally, "ONCOMINE : A Cancer Microarray Database and Integrated Data-mining Platform", *Neoplasia*, Vol. 6, No. 1, pp. 1–6, 2004.

61. Barretina, J., G. Caponigro, N. Stransky, K. Venkatesan, A. a. Margolin, S. Kim, C. J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F. a. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel and L. a. Garraway, "The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity.", *Nature*, Vol. 483, No. 7391, pp. 603–607, 2012.

62. Bernstein, B. E., J. a. Stamatoyannopoulos, J. F. Costello, B. Ren, A. Milosavljevic, A. Meissner, M. Kellis, M. a. Marra, A. L. Beaudet, J. R. Ecker, P. J. Farnham, M. Hirst, E. S. Lander, T. S. Mikkelsen and J. a. Thomson, "The NIH Roadmap Epigenomics Mapping Consortium.", *Nature Biotechnology*, Vol. 28, No. 10, pp. 1045–1048, 2010.

63. Han, J., P. Kraft, H. Nan, Q. Guo, C. Chen, A. Qureshi, S. E. Hankinson, F. B. Hu, D. L. Duffy, Z. Z. Zhao, N. G. Martin, G. W. Montgomery, N. K. Hayward, G. Thomas, R. N. Hoover, S. Chanock and D. J. Hunter, "A Genome-wide Association Study Identifies Novel Alleles Associated with Hair Color and Skin Pigmentation.", *PLoS Genetics*, Vol. 4, No. 5, 2008.

64. Sulem, P., D. F. Gudbjartsson, S. N. Stacey, A. Helgason, T. Rafnar, K. P. Magnusson, A. Manolescu, A. Karason, A. Palsson, G. Thorleifsson, M. Jakobsdottir, S. Steinberg, S. Pálsson, F. Jonasson, B. Sigurgeirsson, K. Thorisdottir, R. Ragnarsson, K. R. Benediktsdottir, K. K. Aben, L. a. Kiemeney, J. H. Olafsson, J. Gulcher, A. Kong, U. Thorsteinsdottir and K. Stefansson, "Genetic Determinants of Hair, Eye and Skin Pigmentation in Europeans.", *Nature Genetics*, Vol. 39, No. 12, pp. 1443–1452, 2007.

65. Natkunam, Y., D. Ph, R. A. Warnke, K. Montgomery, B. Falini and M. V. D. Rijn, "Analysis of MUM1 / IRF4 Protein Expression Using Tissue Microarrays and Immunohistochemistry", *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc*, Vol. 14, No. 7, pp. 686–694, 2001.

66. Sundram, U., J. D. Harvell, R. V. Rouse and Y. Natkunam, "Expression of the B-cell Proliferation Marker MUM1 by Melanocytic Lesions and Comparison with S100, gp100 (HMB45), and MelanA.", *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc*, Vol. 16, No. 8, pp. 802–810, 2003.

67. Duffy, D. L., M. M. Iles, D. Glass, G. Zhu, J. H. Barrett, V. Höiom, Z. Z. Zhao, R. a. Sturm, N. Soranzo, C. Hammond, M. Kvaskoff, D. C. Whiteman, M. Mangino,

J. Hansson, J. a. Newton-Bishop, V. Bataille, N. K. Hayward, N. G. Martin, D. T. Bishop, T. D. Spector and G. W. Montgomery, "IRF4 Variants Have Age-specific Effects on Nevus Count and Predispose to Melanoma.", *American Journal of Human Genetics*, Vol. 87, No. 1, pp. 6–16, 2010.

68. Peña Chilet, M., M. Ibarrola-Villava, M. Martin-González, M. Feito, C. Gomez-Fernandez, D. Planelles, G. Carretero, A. Lluch, E. Nagore and G. Ribas, "rs12512631 on the Group Specific Complement (Vitamin D-binding Protein GC) Implicated in Melanoma Susceptibility.", *PloS One*, Vol. 8, No. 3, p. e59607, 2013.

69. Cai, Y.-H. and H. Huang, "Advances in the Study of Protein-DNA Interaction. ", *Amino Acids*, Vol. 43, No. 3, pp. 1141–1146, 2012.

70. Orlando, V., "Mapping Chromosomal Proteins in Vivo by Formaldehyde- Immunoprecipitation", *Elsevier Science*, Vol. 0004, pp. 99–104, 2000.

71. Fullwood, M. J. and Y. Ruan, "ChIP-based Methods for the Identification of Long-range Chromatin Interactions.", *Journal of Cellular Biochemistry*, Vol. 107, No. 1, pp. 30–39, 2009.

72. Oliphant, A. R., C. J. Brandl and K. Struhl, "Defining the Sequence Specificity of DNA-Binding Proteins by Selecting Binding Sites from Random-Sequence Oligonucleotides : Analysis of Yeast GCN4 Protein", *Molecular and Cellular Biology*, Vol. 9, No. 7, pp. 2944–2949, 1989.

73. Ferna, M., T. S. Chen and A. Y. Ting, "Protein - Protein Interaction Detection in Vitro and in Cells by Proximity Biotinylation", *Journal of American Chemical Society*, , No. 130, pp. 9251–9253, 2008.

74. Sanger, F. and S. Nicklen, "DNA Sequencing with Chain-terminating", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 74, No. 12, pp. 5463–5467, 1977.

75. Smith, M. C. M., A. Mountain and S. Baumherg, "Sequence Analysis of the Bacillus subtilis argC Promoter Region", *Gene*, Vol. 49, pp. 53–60, 1986.

76. Morozova, O. and M. a. Marra, "Applications of Next-generation Sequencing Technologies in Functional Genomics.", *Genomics*, Vol. 92, No. 5, pp. 255–264, 2008.

77. Loman, N. J., R. V. Misra, T. J. Dallman, C. Constantinidou, S. E. Gharbia, J. Wain and M. J. Pallen, "Performance Comparison of Benchtop High-throughput Sequencing Platforms.", *Nature Biotechnology*, Vol. 30, No. 5, pp. 434–439, 2012.

78. Metzker, M. L., "Sequencing Technologies - The Next Generation.", *Nature Reviews. Genetics*, Vol. 11, No. 1, pp. 31–46, 2010.

79. Clarke, J., H.-C. Wu, L. Jayasinghe, A. Patel, S. Reid and H. Bayley, "Continuous Base Identification for Single-Molecule Nanopore DNA Sequencing.", *Nature Nanotechnology*, Vol. 4, No. 4, pp. 265–270, 2009.

80. Pennisi, E., "Search for Pore-fection", *Science*, Vol. 336, pp. 534–537, 2012.

81. Johnson, D. S., A. Mortazavi, R. M. Myers and B. Wold, "Genome-wide Mapping of in Vivo Protein-DNA Interactions", *Science*, Vol. 316, No. June, pp. 1497–1503, 2007.

82. Ren, B., J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell and R. A. Young, "Genome-Wide Location and Function of DNA Binding Proteins", *Science*, Vol. 290, pp. 230Machanick20116–2310, 2000.

83. Rhee, H. S. and B. F. Pugh, "Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-nucleotide Resolution.", *Cell*, Vol. 147, No. 6, pp. 1408–1419, 2011.

84. Raney, B. J., M. S. Cline, K. R. Rosenbloom, T. R. Dreszer, K. Learned, G. P.

Barber, L. R. Meyer, C. a. Sloan, V. S. Malladi, K. M. Roskin, B. B. Suh, A. S. Hinrichs, H. Clawson, A. S. Zweig, V. Kirkup, P. a. Fujita, B. Rhead, K. E. Smith, A. Pohl, R. M. Kuhn, D. Karolchik, D. Haussler and W. J. Kent, "ENCODE Whole-genome Data in the UCSC Genome Browser (2011 Update).", *Nucleic Acids Research*, Vol. 39, No. Database issue, pp. D871–875, 2011.

85. Robertson, G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O. L. Griffith, A. He, M. Marra, M. Snyder and S. Jones, "Genome-wide Profiles of STAT1 DNA Association Using Chromatin Immunoprecipitation and Massively Parallel Sequencing.", *Nature Methods*, Vol. 4, No. 8, pp. 651–657, 2007.

86. Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg, "Ultrafast and Memory-efficient Alignment of Short DNA Sequences to the Human Genome.", *Genome Biology*, Vol. 10, No. 3, p. R25, 2009.

87. Mendoza-Parra, M.-A., W. Van Gool, M. A. Mohamed Saleem, D. G. Ceschin and H. Gronemeyer, "A Quality Control System for Profiles Obtained by ChIP Sequencing.", *Nucleic Acids Research*, Vol. 41, No. 21, p. e196, 2013.

88. Zhang, Y., T. Liu, C. a. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li and X. S. Liu, "Model-based Analysis of ChIP-Seq (MACS).", *Genome Biology*, Vol. 9, No. 9, p. R137, 2008.

89. Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass, "Simple Combinations of Lineage-determining Transcription Factors Prime Cis-regulatory Elements Required for Macrophage and B Cell Identities.", *Molecular Cell*, Vol. 38, No. 4, pp. 576–89, 2010.

90. Spyrou, C., R. Stark, A. G. Lynch and S. Tavaré, "BayesPeak: Bayesian Analysis of ChIP-seq Data.", *BMC Bioinformatics*, Vol. 10, p. 299, 2009.

91. Cairns, J., C. Spyrou, R. Stark, M. L. Smith, A. G. Lynch and S. Tavaré,

"BayesPeak–an R Package for Analysing ChIP-seq Data.", *Bioinformatics*, Vol. 27, No. 5, pp. 713–714, 2011.

92. Bailey, T. L., M. Boden, F. a. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li and W. S. Noble, "MEME SUITE: Tools for Motif Discovery and Searching.", *Nucleic Acids Research*, Vol. 37, pp. 202–208, 2009.

93. Machanick, P. and T. L. Bailey, "MEME-ChIP: Motif Analysis of Large DNA Datasets.", *Bioinformatics*, Vol. 27, No. 12, p. 1696, 2011.

94. McLean, C. Y., D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger and G. Bejerano, "GREAT Improves Functional Interpretation of Cis-Regulatory Regions.", *Nature Biotechnology*, Vol. 28, No. 5, pp. 495–501, 2010.

95. Eden, E., R. Navon, I. Steinfeld, D. Lipson and Z. Yakhini, "GOrilla: A Tool for Discovery and Visualization of Enriched GO Terms in Ranked Gene Lists.", *BMC Bioinformatics*, Vol. 10, p. 48, 2009.

96. Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee and B. L. Ebert, "Gene Set Enrichment Analysis : A Knowledge-based Approach for Interpreting Genome-wide", *Proceedings of the National Academy of Sciences of the United States of America*, 2005.

97. Kumar, C. M., U. Sathisha, S. Dharmesh, A. A. Rao and S. A. Singh, "Interaction of Sesamol (3,4-methylenedioxyphenol) with Tyrosinase and its Effect on Melanin Synthesis", *Biochimie*, Vol. 93, No. 3, pp. 562 – 569, 2011.

98. Mitra, D. and D. E. Fisher, "Transcriptional Regulation in Melanoma", *Hematology/Oncology Clinics of North America*, Vol. 23, No. 3, pp. 447 – 465, 2009, advances in Melanoma.

99. MustafaCan, A., *Characterization of IRF4 in Melanoma Cell Lines*, Master's Thesis, Bogazici University, 2014.