Long Range Interactions of the IRF4 Promoter in Myeloma and Melanoma

Kristján Hólm Grétarsson

Ritgerð til meistaragráðu
Háskóli Íslands
Læknadeild
Heilbrigðisvísindasvið

HÁSKÓLI ÍSLANDS

Langdrægar tengingar IRF4 stýrils í mergæxlum og sortuæxlum

Kristján Hólm Grétarsson

Ritgerð til meistaragráðu í líf- og læknavísindum

Umsjónarkennari: Eiríkur Steingrímsson

Leiðbeinandi: Erna Magnúsdóttir

Aðrir í Meistaranámsnefnd: Stefán Þórarinn Sigurðsson

Læknadeild

Heilbrigðisvísindasvið Háskóla Íslands

Júní 2016

Long Range Interactions of the IRF4 promoter in Myeloma and Melanoma

Kristján Hólm Grétarsson

Thesis for the degree of Master of Science

Supervisor: Eiríkur Steingrímsson

Advisor: Erna Magnúsdóttir

Masters committee: Stefán Þórarinn Sigurðsson

Faculty of Medicine

School of Health Sciences

June 2016

Prentun: Háskólaprent efh.

Reykjavík, 2016

Ágrip

Skipulag erfðamenginsins og lögun litninga hefur áhrif á tjáningu gena. Skipulagið er breytilegt milli frumna, þroskunarstigs og í frumuhringnum og hefur mikil áhrif á tengingar milli efliraða og stýrla og hefur þar með áhrif á genatjáningu. Það er því áhugavert að skoða langdrægar tengingar út frá stýrli gens til að rannsaka stjórnun þess. Í B frumum stjórnar umritunarþátturinn IRF4 flokkaskiptri endurröðun og hefur lykiláhrif á þroskun B frumna úr kímstöðvum í mótefnisseytandi plasma frumur. IRF4 er einnig nauðsynlegt fyrir lifun mergæxlisfrumna og gegnir auk þess hlutverki í litfrumum og sortuæxlum.

Markmið þessarar rannsóknar var að mæla langdrægar tengingar út frá IRF4 stýrlinum með 3C tækninni sem byggir á hremmingu á stellingu litninga og bera saman milli mergæxla- og sortuæxlis-frumulína til að finna frumusértækar efliraðir sem hafa áhrif á tjáningu IRF4.

Til að gera þetta var 3C tæknin sett upp og má nú nota til að mæla langdrægar tengingar. 3C tækninni var beitt á fimm frumulínur; Waldenström macroglobulinemia frumulínuna RPCI-WM1, mergæxla frumlínuna U266B1, sortuæxlis frumulínurnar 501mel og SKmel28 og HEK293T frumulínuna sem neikvætt viðmið í samanburðinum. Þetta leiddi í ljós að 17kb svæði ofan við IRF4 stýrilinn myndaði B frumulínu-háða litnislykkju yfir á IRF4 stýrilinn. Sviperfða gögn úr ENCODE verkefninu voru notuð til að rannsaka þetta 17kb svæði nánar og fundust tvær mögulegar efliraðir innan svæðisins. Með litnis-mótefnaútfellingu á histón umbreytingum og IRF4 fundust merki um efliraðir í svæðunum tveimur í RPCI-WM1 frumulínunni. Að lokum var efliraða-virkni annars svæðisins innan B frumulínu-háðu litnislykkjunnar yfir á IRF4 stýrilinn prófuð . Þar kom í ljós að svæðið hefur daufa efliraðavirkni.

Abstract

Gene expression is heavily dependent on the conformation of the genome as genetic elements loop out to interact with each other. Chromosome organization is highly dynamic, varying during the cell cycle, developmental stage and between different cell types affecting enhancer promoter connections, consequently affecting the corresponding gene expression. It is interesting to look at differences in long range chromatin interactions to study the underlying control of gene expression. In the B-lymphocyte lineage, expression of the transcription factor IRF4 leads to B-cell heavy chain class switch recombination and the generation of plasma cells from germinal centre B lymphocytes and is required for the survival of myeloma cell. Additionally, IRF4 plays a role in melanoma and pigmentation.

The goal of this project was to look at the long range interactions of elements responsible for the transcription of the IRF4 gene in myeloma and melanoma using the chromosome conformation capture (3C) technique to study chromatin interaction.

The chromosome conformation capture (3C) technique has been set up and can be used to find long range interactions. The 3C method was applied to five cell lines, the Waldenström macroglobulinemia cell line RPCI-WM1, the multiple myeloma cell line U266B1, the melanoma cell lines 501mel and SKmel28 and the human embryonic kidney cell line 293T (HEK293T). This method revealed a 17kb region upstream of the IRF4 promoter that forms a strong B cell linage-specific chromatin loop to the IRF4 promoter. Epigenetic data from the ENCODE project was utilized to study the 17kb looping region and identified two potential enhancer regions. The enhancer indicators of the regions were validated in the RPCI-WM1 cells using histone modification and IRF4 chromatin immunoprecipitation. Finally the ability of the more promising sub-sequence within the B cell lineage-specific looping region in the IRF4 locus, was tested in a transcription activation assay, revealing a modest enhancer activity of the region.

Acknowledgements

I would like to thank my advisor, Dr. Erna Magnúsdóttir, for the opportunity to work on this exciting project and for her support and guidance during the project. Professor Eiríkur Steingrímsson for allowing me to join his lab at Læknagarður in 2014 and both Erna and Eiríkur for their encouragements and scientific teaching.

I would like to thank my masters committee Erna, Eiríkur and Dr. Stefán Þórarinn Sigurðsson for their support, proofreading and advice on my thesis.

Further, I would like to thank current and former members of the laboratory of Professor Eiríkur Steingrímsson; Margrét Helga Ögmundsdóttir, Katrín Möller, Josue Ballesteros, Dilixiati Remina Diahann Atacho, Kristín Bergsteinsdóttir, Sara Sigurbjörnsdóttir, Valerie Fock, Óskar Örn Hálfdánarson, Ásgeir Örn Arnþórsson and Sigurður Rúnar Guðmundsson and of the laboratory of Erna Magnúsdóttir; Aðalheiður Elín Lárusdóttir, Kimberley Anderson and Klara Hansdóttir.

Áslaug Jónasdóttir and Aðalbjörg Jónasdóttir at deCODE get thanks for teaching and allowing me to use the droplet digital PCR setup at deCODE.

Finally, I especially want to thank my family and friends for their invaluable support during those two years of studies.

Table of Contents

Figures

Tables

Abbreviations

3C	Chromosome conformation capture
BAC	Bacterial artificial chromosome
BCR	B cell receptor
bp	Base pair
BSA	Bovine serum albumin
ChIA-PET	Chromatin Interaction analysis by paired-end tag sequencing
ChIP	Chromatin immunoprecipitation
CSR	Class switch recombination
CT	Chromosome territories
ddPCR	droplet digital PCR
DMEM	Dulbecco´s Modified Eagle Medium
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
ENCODE	the Encyclopedia of DNA Elements
eRNA	Enhancer RNA
FBS	Fetal bovine serum
H3K27ac	Histone 3 lysine 27 acetylation
H3K4me1	Histone 3 lysine 4 monomethylation
H3K4me3	Histone 3 lysine 4 trimethylation
IRF4	Interferon regulatory factor 4
kb	Kilo base
MPRA	Massive parallel reporter assays
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PMSF	Phenylmethane sulfonyl fluoride
qPCR	quantitative PCR
RNA	Ribonucleic acid
RPMI	Roswell Park Memorial Institute
SDS	Sodium dodecyl sulfate
SHM	Somatic hypermutation
TAD	Topologically associated domains
WE	Weak Enhancer

1. Introduction

1.1 B-lymphocyte and Plasma Cell Development

B-lymphocytes develop in the bone marrow from haematopoietic stem cells. In this process the immungoglobulin heavy-chain gene segments are rearranged, and successful B-lymphocyte clones undergo further development including clonal expansion and rearrangement of immunoglobulin light-chain gene segments to generate functional B-cell receptors (BCR). After negative selection, cells that interact with host receptors are removed, pre-B-lymphocytes exit the bone marrow and enter the secondary lymphoid tissues (lymph nodes, spleen and mucosal lymphoid tissue) (Shapiro-Shelef & Calame, 2005; Travers et al., 2001). B-lymphocytes continue their development in those tissues where they pass through two transitional stages (T1 and T2) before coming fully mature (Chung et al., 2003). A small fraction of the transitional B-lymphocytes occupy the splenic marginal zone as naive non-circulating marginal zone B-lymphocytes and become the first B-lymphocytes to respond to a foreign antigen by differentiating into plasma cells and therefore provide an important initial response to antigens. However most B-lymphocytes mature into long lived naive follicular B-lymphocytes where they circulate between the secondary lymphoid tissues and the bone marrow until they either die or become activated by antigen interaction. Follicular B-lymphocytes can either respond rapidly like the marginal-zone B-lymphocytes, undergoing plasmacytic differentiation to form extrafollicular foci of short lived and cycling plasmablasts and eventually long lived, plasma cells or more slowly through the germinal centre reactions (Nutt et al., 2015; Shapiro-Shelef & Calame, 2005).

B-lymphocytes migrate to the germinal centres within the lymph nodes after they have been activated by an antigen. There they undergo affinity maturation against the same antigen and the hypervariable regions of the immunoglobin heavy chain locus undergo somatic hypermutations (SHM). The structure of the antibody is further altered by class switch recombination (CSR) which produces antibodies of different immunoglobulin isotypes. During the germinal centre reactions the selection is carried through by the T helper cells. The germinal centre then gives rise to plasma cells with high antigen affinity and can become long lived if they are successful in finding a niche in the bone marrow. Another type of cells that exit the germinal centre are memory B-lymphocytes. Memory B-lymphocytes have high affinity antibodies on their surface (after undergoing SHM and CSR in the germinal centre), they do not secrete antibodies but have the ability to proliferate and differentiate quickly into plasma cells upon secondary exposure to antigen (De Silva & Klein, 2015). To complete the differentiation into mature antibody secreting plasma cells the B-lymphocytes need to turn off proliferation and activate the functions of mature B-lymphocytes; immunoglobulin CSR and SHM, BCR signalling and antigen presentation (Shaffer et al., 2002). Subsequently induction of plasma cell functions is required; expansion of the endoplasmic reticulum, secretion and production of antibodies and survival in unique bone narrow niches (Nutt et al., 2015). These steps are coordinated by expression switch from proteins such as BCL-6 and PAX5, that are crucial for the execution of the B-cell transcriptional programme and functions (Kitano et al., 2011; Schebesta et al., 2007), to expression of IRF4, BLIMP1 and XBP proteins that regulate the plasma cell transcriptional programme (Hu et al., 2009; Morgan et al., 2012; Ochiai et

al., 2013; Shaffer et al., 2002). The same genes are known to be deregulated in various lymphomas and/or myelomas (Carrasco et al., 2007; Parekh et al., 2007; Pasqualucci et al., 2006; Zheng et al., 2013).

1.2 The Transcription Factor IRF4

Interferon regulatory factor 4 (IRF4) is a transcription factor belonging to the superfamily of interferon regulator factors (IRFs). IRF4 binds DNA weakly on its own and needs to bind DNA co-operatively with other transcription factors, namely IRF8, PU.1 and Spi-B to carry out its role (Recaldin & Fear, 2016). Mice lacking IRF4 revealed that IRF4 is dispensable during early B lymphocyte development but has an important role in late B lymphocyte development as the mice did not form germinal centres (GCs) or plasma cells and serum immunoglobulin (Ig) levels were severely reduced (De Silva et al., 2012; Mittrucker et al., 1997).

Analysis of IRF4 expression has revealed high protein expression in plasma cells but most germinal centre B lymphocytes lack IRF4 except a small proportion of B cells within the light zones of germinal centres. Those IRF4 positive B lymphocytes have plasmablastic features and most of them co-express B lymphocyte-induced maturation protein 1 (BLIMP1). BLIMP1 is the master regulator of plasma cell differentiation suggesting that these cells are plasma cell precursors (Recaldin & Fear, 2016; Shapiro-Shelef et al., 2003). IRF4 has an essential role in class switch recombination (CSR) and a possible role in somatic hypermutations (SHM) as IRF4-deficient mice fail to induce activation-induced cytidine deaminase (AID) expression which is needed for CSR and SHM (Klein et al., 2006; Ochiai et al., 2013; Recaldin & Fear, 2016). It is now known that IRF4 determines two different cell fate transitions in mature B-lymphocyte development, the entry of antigen activated B-lymphocytes into the germinal centre reactions and the formation of plasma cells from germinal centre B lymphocytes. The underlying mechanism for this context dependent role of IRF4 might stem from the difference in expression levels, as IRF4 is expressed at low levels in naive B cells but is up-regulated during PC differentiation (Ochiai et al., 2013; Sciammas et al., 2006; Willis et al., 2014). It is also possible that the affinity of the membrane bound antibody for antigen determines the level of IRF4 expression and IRF4 DNA motif recognizing ability is dependent upon IRF4 concentration. Initially the B cell receptor shows only weak antigen affinity and the B lymphocyte enters the germinal centre reaction as low amounts of IRF4 activates BCL6 expression through binding to ETS-IRF or AP-1-IRF motifs. As the germinal centre reaction continues and the B lymphocytes undergo somatic hypermutations the antigen affinity increases, leading to more IRF4 expression. This eventually causes BLIMP1 expression through binding of IRF4 as a homodimer to interferon sequence response elements (ISREs) near the BLIMP1 promoter and escape from the germinal centre reactions and plasma cell differentiation. (Ochiai et al., 2013; Recaldin & Fear, 2016; Willis et al., 2014). The function of IRF4 to drive B cell to plasma cell differentiation is also seen immediately after follicular B cell activation as a double-negative feedback loop between IRF4 and IRF8 regulates the developmental trajectories of B cells. B cells expressing high amount of IRF4 undergo rapid extrafollicular plasmablast differentiation compared to B cells expressing high amount of IRF8 that undertake a slower germinal centre response (Xu et al., 2015). This is consistent with the idea that high concentration of IRF4 pushes B cells toward plasma cell differentiation.

1.3 Multiple Myeloma and Waldenström Macroglobulinemia

Multiple myeloma is the cancer of malignant plasma cells. Multiple myeloma is incurable, but treatable with 18 new cases per year and a 60% (male) and 49% (female) 5 year survival rate in Iceland (Jónasson & Tryggvadóttir, 2012). Distinct clinical phases of myeloma can be recognized such as monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM). Even though these phases lack the clinical symptoms of myeloma, such as anemia, bone pain, organ failure and infections they have similar genetic features of symptomatic myeloma (Hanahan & Weinberg, 2011).

Waldenström macroglobulinemia is a rare, indolent hematologic malignancy thought to be of post-germinal center B-lymphocytes origin. These cells have undergone somatic hypermutations but are transformed before class switch recombination. The cells secrete monoclonal IgM, leading to increased viscosity of the blood causing some of the symptoms associated with Waldenström macroglobulinemia (Chitta et al., 2013; Walsh et al., 2005).

IRF4 is essential for the survival of myeloma cells. Loss of IRF4 in myeloma cells affects the cells in multiple ways through its master regulating influences on metabolic control, membrane biogenesis, cell cycle progression, cell death, transcriptional regulation and plasmacytic differentiation. Eventually, these disruptions lead to cell death, demonstrating the role of IRF4 as a master regulator of a malignancy specific gene expression programme irrespective of their molecular subtype (Shaffer et al., 2008). The dependency of myeloma cells on IRF4 can be described as 'non-oncogene addiction' as the dependency on IRF4 is independent of mutations or locus alterations of IRF4. An aberrant IRF4 regulation network reminiscent of fused mature B cell and plasma cell IRF4 regulation network was uncovered when comparing the regulation network of IRF4 between normal plasma cells and myeloma cells (Shaffer et al., 2008; Solimini et al., 2007).

1.4 Melanoma and IRF4

IRF4 expression is also found in a subset of malignant melanomas and melanocytes (Grossman et al., 1996; Natkunam et al., 2001). Melanoma is the cancer of melanocytes which are melanin producing cells mainly located in the epidermis, the iris and hair follicles but have been also found in the inner ear, the nervous system and the heart (Cichorek et al., 2013). The number of new cases of melanoma per year is 40-45 in Iceland with 85% (male) and 95% (female) 5 year survival rate (Jónasson & Tryggvadóttir, 2012). The role of IRF4 in melanoma and melanocytes is less clear than in plasma cell differentiation and hematolymphoid malignancies. However variation in the IRF4 locus has been associated with pigmentation polymorphism in humans (Praetorius et al., 2013; Sulem et al., 2007) and a recent study demonstrated a direct role of IRF4 in pigmentation (Praetorius et al., 2013). IRF4 expression is regulated by a melanocyte specific enhancer located on intron 4 of IRF4 (Praetorius et al., 2013; Thurman et al., 2012; Visser et al., 2015). The melanocyte master regulator MITF, along with the transcription factor activator protein 2a (TFAP2A) bind the intron 4 enhancer on IRF4 and activate IRF4 expression. IRF4 along with MITF in turn activate the expression of the gene encoding the pigmentation enzyme Tyronsinase (TYR) (Praetorius et al., 2013).

Interestingly, MITF, which is also expressed in naïve B cells, has an antagonistic effect on IRF4 expression in those cells as IRF4 is repressed. MITF therefore suppresses spontaneous plasma cell differentiation in naïve B cells. MITF most likely does not bind directly to the promoter of IRF4 which might suggest a repression through binding to a nearby regulatory element or an indirect repression for example inhibiting NF-βK signaling upstream of IRF4 expression (Lin et al., 2004). However, it has been shown that the melanocyte enhancer at intron 4 of IRF4 has a repressing effect on IRF4 transcription through TFAP2A binding to the same region in B lymphoma cells (Do et al., 2010). Even though MITF has not been directly linked to repress IRF4 through this region in B lymphocytes or lymphoma, given its role in IRF4 regulation in melanocytes MITF might be having an antagonistic effect on IRF4 expression through the same regulatory region in lymphocytes and melanocytes.

1.5 Enhancers

DNA cis-regulatory elements are important for tissue specific gene expression. Genome-wide genetic association studies have suggested that about 85% of disease-associated variants lie outside protein-coding regions, highlighting the importance of understanding regulatory elements (Hindorff et al., 2009; Kheradpour et al., 2013).

Enhancers are DNA elements that can stimulate promoter activity when they come in direct contact with them through chromatin looping, irrespective of sequence orientation (van Arensbergen et al., 2014). Enhancers are typically few hundred base pairs and contain binding sites for transcription factors. Upon binding of multiple transcription factors, including lineage specific transcription factors and sequence-dependent effectors of signaling pathways, the enhancers become activated. In turn, they affect their target genes through chromatin looping to their corresponding promoters (Calo & Wysocka, 2013). To initiate transcription, some transcription factors are dependent upon a set of proteins called coactivators. Coactivators lack sequence-specific DNA binding domains but are cooperatively recruited by transcription factors to promoters and enhancers. Coactivators function either as histone modifiers, ATP dependent chromatin remodelers or mediators of long-range contacts with the transcriptional machinery at promoters (Calo & Wysocka, 2013). One common coactivator that might have a role in pre-marking of tissue-specific enhancers is P300/CBP which is recruited to enhancers through distinct transcription factor classes. P300/CBP is a histone acetyltransferase (HAT) and one of its main substrates is H3K27 but H3K27ac distinguishes active enhancer states from inactivated enhancers or those poised for activation (Calo & Wysocka, 2013; Q. Jin et al., 2011). In concordance, P300 binding has been shown to result in good prediction of enhancer location in the genome (Visel et al., 2009).

Enhancers can affect multiple promoters and distinct enhancers can affect a single promoter in a tissue specific fashion depending on the TFs that bind to them (Hardison & Taylor, 2012). Enhancers can act over large genomic distances to affect promoters, sometimes over hundreds of kb and although rare, between chromosomes (trans) (Lomvardas et al., 2006; van Arensbergen et al., 2014). This has been suggested by multiple different approaches attempting to identify linked enhancer and promoter loci: direct contact between enhancer and promoter from 3C-based experiments (Sanyal et al., 2012), correlative DNase hypersensitivity sites at promoter and distant loci (Thurman et al., 2012), correlative chromatin state of enhancer and promoter loci and correlative enhancer RNA expression to promoter activity (Andersson et al., 2014; van Arensbergen et al., 2014). But the mere distance between

enhancers and the promoter is not sufficient to describe functional interaction as enhancers often ignore the nearest promoters to interact with more distant promoters and recent study suggest that only 7% of enhancers interact with its nearest promoter (Sanyal et al., 2012).

Recently, a wide variety of high-throughput technologies have been designed, termed massive parallel reporter assays (MPRA) (Arnold et al., 2013; Birnbaum et al., 2014; Kwasnieski et al., 2014). With the purpose of simultaneously studying the effects of nucleotide variation within an enhancer on the activity of the enhancer and mapping functional transcription factor binding sites at single-nucleotide resolution. Gaining information on the basis of cell specific activity of the enhancers (Inoue & Ahituv, 2015). The majority of single nucleotide variants in these enhancers had modest effects on transcriptional activation, suggesting that enhancers are highly robust to single nucleotide changes. Also, there was a correlation between the magnitude of functional impact and the location of predicted transcription factor binding sites such that nucleotides located within transcriptional binding sites had more effect on the enhancer activity than other nucleotides. .

Another study reported a large impact of the majority (86%) of single-nucleotide substitutions in a conserved enhancer on expression, attributing this to affinity changes of transcription factor binding sites, gain and loss of binding sites and transcription factor competition (Kwasnieski et al., 2012). A more recent study also reported a significant impact of a smaller proportion of enhancer nucleotides (11%) on enhancer activity. Moreover, those same bases also tended to cluster within the enhancer and to be evolutionary conserved (S. Li & Ovcharenko, 2015).

Currently there are no absolute rules known for the identification of regulatory elements other than testing individual sequences with plasmid-based reporter assays (Dogan et al., 2015; Inoue & Ahituv, 2015). Reporter assays for enhancer activity are based on cloning the sequence of interest into a reporter vector containing a common or tissue specific promoter in front of a reporter gene, e.g. luciferase, lacZ or GFP. The activity of the enhancer is then tested by measuring the expression of the reporter gene in cell lines or transgenic animals (Inoue & Ahituv, 2015). Many attempts have been made to predict enhancers based on clustered TF motifs, TF ChIP-Seq data, epigenetics markers and non-coding evolutionary conserved regions (Hardison & Taylor, 2012). The histone code, the specific combinations of covalent histone modifications, has been speculated to control the role of the underlying DNA sequence (Kwasnieski et al., 2014). In a recent effort, the ENCODE project attempted to classify DNA segments based on covalent histone modifications by ChIP-seq along with other functional genomic data such as DNase-seq, FAIRE-seq, RNA polymerase II and CTCF ChIP-seq in multiple cell lines (Consortium, 2012; Kwasnieski et al., 2014). Of the seven major classes of genome states discovered, two were considered likely to be enhancers. Both classes are described as regions of open chromatin associated with H3K4me1 signal, enriched for TFs and coactivators known to act at enhancers (including EP300) and having an excess of RNA elements without poly-A tails and methyl-cap RNA, hinting towards RNA transcription at active enhancers. Of these two states one was termed weak enhancer (WE) based on lower signaling strength of H3K27ac and H3K4me1 histone modifications (Consortium, 2012). These epigenetics features had previously been successful in enhancer prediction (Heintzman et al., 2009; Rada-Iglesias et al., 2011; Visel et al., 2009). However those features only show modest estimate for active elements when tested by MPRA in cell lines (26%)

(Kwasnieski et al., 2014). One study reported higher accuracy of key transcription factor occupancy for enhancer identification. By using genome-wide measurement of epigenetic features, such as histone modifications and occupancy by transcription factors they attempted to quantify the contribution of each factor in enhancer prediction. They found that DNA segments bound by key transcription factors, the coactivator P300, SMAD, had the histone modifications of active enhancer and located on DNase hypersensitive sites correctly predicted enhancer activity in the majority of cases (80%) by use of reporter systems. However DNA segments containing chromatin with the very same histone modifications but with no binding by the key transcription factors were almost never validated as enhancers experimentally (Dogan et al., 2015).

Interestingly, active enhancers appear to be bound by general transcription factors (GTF) and RNA polymerase II (PolII) forming the preinitiation complex (PIC), causing the enhancer to be transcribed and producing non-coding RNA termed enhancer RNA (eRNA). The eRNA transcription appears to have a functional role in regulation of gene expression as knock down of at least some eRNAs causes reduction in mRNA levels of nearby genes (Lam et al., 2013; Melo et al., 2013). In the context of enhancer identification, especially finding highly active enhancers, eRNA production may be used as enhancer transcription correlates well with enhancer activity (Calo & Wysocka, 2013; Natoli & Andrau, 2012).

1.6 Genome Organization

The packaging of DNA into the eukaryotic nucleus is a tremendous structural challenge and is governed by rules related to genomic function. In order to understand how the genome functions it is not enough to study the information encoded in the DNA sequence. It is also important to explore how the genome is organized in three dimensional space to reveal important topological information as the third dimensional conformation of the genome is related to expression pattern (Gibcus & Dekker, 2013). Certain parts of the genome are differently expressed depending on the cell type, cell cycle and developmental status and the environment. This difference in expression affects the folding of the genome (Heidari et al., 2014).

Looking at the genome organization by investigating the radial placement of chromosomes and genes within the nucleus, much can be learned on how the chromosomes are arranged. Because of their large size, individual chromosomes occupy well defined space termed chromosome territories (CT). Interactions of loci in trans (between chromosomes) are much rarer than interactions in cis (within chromosomes) demonstrating the existence of chromosomal territories (Gibcus & Dekker, 2013). In addition, more connections or intermingling between neighboring chromosome territories have been shown using high-resolution in situ hybridization (Branco & Pombo, 2006) and Hi-C (Zhang et al., 2012). CTs are not randomly located in the nucleus and it has been noted that gene poor chromosomes are more likely to be near the nuclear periphery and gene rich and smaller chromosomes (16, 17, 19, 20, 21, 22) preferentially interact with each other and co-localize in the center of the nucleus (de Wit & de Laat, 2012; Lieberman-Aiden et al., 2009). The position distribution of genes and chromosomes within the nucleus is characteristic for each chromosome and gene and is tissue specific. This distribution is a statistical property and may be highly variable between individual cells. The precise position of a gene within the nucleus is not sufficient to determine the activity of the gene, however it’s positioning along

the nuclear envelope can affects a gene's activity. The nuclear periphery is enriched in condensed heterochromatin and lamina-associated domains in the genome have been identified. They are typically 0.1-1 Mb in size, gene poor and contain genes which are either silenced or expressed at low levels. This suggests that the periphery is a transcriptionally repressive environment (Dekker & Misteli, 2015). Spatial gene clustering also has an impact on 3D genome organization. Efficiency of transcription and RNA processing is increased through clustering of co-regulated genes at sites known as transcription factories. Transcription factories are subnuclear foci composed of active RNAPII and other transcriptional accessory and regulatory factors (Deng et al., 2013).

1.7 Chromosome Conformation Capture (3C) based techniques

Over the last decade the chromosome conformation capture (3C) techniques have revolutionized studies regarding genome organization. These methods are based on ligating restriction enzyme digested distal genomic fragments that co-localize in space via chromosome fibre interactions. The 3D organization of the genome is fixed in point with a fixation agent. The fixed chromatin is digested with a restriction enzyme and the (sticky) ends of the cross-linked fragments are ligated under conditions that promote intramolecular ligations. The frequency of these ligated fragments, which reflect the interaction between two distant genomic loci, are then studied with quantitative methods (such as qPCR or droplet digital PCR) to measure the number of ligation events, using primers located near the ligation junctions (Hagege et al., 2007).

The 3C technique is limited in that it only allows for the detection of one ligation product at a time (one on one). However, multiple modifications have been made to the original 3C technique (Dekker et al., 2002) by combining it with next generation sequencing. These include the circular chromosome conformation capture (4C) (Zhao et al., 2006) that utilizes inversion PCR to enable the study of interactions of all interactions too one anchor element (one vs all) and chromosome conformation capture carbon copy (5C) (Dostie et al., 2006), a method to study all interaction within a defined region, often up to few megabases (limited all vs all) (de Wit & de Laat, 2012). The HiC method (Lieberman-Aiden et al., 2009) attempts to expand the 3C method by theoretically catching all interacting loci within the genome at the cost of a lower resolution, only capturing the most frequent interactions. HiC can be coupled with immunoprecipitation of a certain protein to only study interactions that associate with a specific protein with a method named Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChiA-PET) (de Wit & de Laat, 2012). Current methods can even enrich for sequence specific ligation products (i.e. promoters) to increase resolution while lowering sequencing coverage required (Capture Hi-C) (Mifsud et al., 2015) and the updated version of HiC, termed in-situ HiC claims it can obtain 1kb resolution (Rao et al., 2014).

1.8 Topologically Associated Domains

The Hi-C technique is a powerful tool that enables the complete, but cell averaged, genome structure to be studied and recent improvements in HiC, sequencing power and data analysis have given us a lot of information about the genome organization and how it comes to be (de Wit & de Laat, 2012). It has been suggested multiple times that the chromosomes form highly self-interacting regions

termed topologically associated domains (TAD) that represent boundary regions in the genome (de Wit & de Laat, 2012; Dixon et al., 2012; G. Li et al., 2012). With HiC 2,200 topological domains in mouse embryonic stem cells (mESCs) have been identified with a median size of 880kb that occupy ~91% of the genome (Dixon et al., 2012). More recent studies, using kilobase resolution based on in-situ HiC observed domains ranged in size from 40 kb to 3 Mb (median size 185 kb) and using ChIA-PET against CTCF found domains ranged in 20kb to 2 Mb respectively (F. Jin et al., 2013; Tang et al., 2015). With improved techniques, the size of highly intra-relating loci within a region, TAD, is getting smaller. It should be noted that TADs are sometimes hard to define, borders between TADs can be unclear and smaller domains with higher contact frequency can be seen within TADs as a consequence of the complexity of Hi-C contact maps and/or dynamic TAD formation (Rowley & Corces, 2016).

TADs can be thought of as fundamental units in genome organization and are responsible for directing regulatory elements towards their appropriate promoters (Bickmore, 2013). A model based on the fact that different cell types share most of their topological domains and most of the difference (dynamic interactions) is within those topological domains, states that higher chromatin order is mostly stable between cell types but regions within the domains interact in more cell type specific manner and regulate cell type specific events (Dixon et al., 2012). In agreement with this a great conservativism has been seen when comparing domains between cell types and even organisms. Likewise, syntenic regions in mouse and human cell lines shared significant part of their domains, indicating conservation of higher order chromatin structure (Dixon et al., 2012; Rao et al., 2014). The boundary regions between the topological domains show strong enrichment for the insulator protein CTCF, transcription start sites, housekeeping genes and actively transcribed genes (Dixon et al., 2012). An interesting phenomenon based on the comparison of domains in pluripotent and differentiated cells is the fact that domain boundaries, that show similar distribution in both cell lines, seem to predetermine the spread of heterochromatin in pluripotent cells; the chromatin modification marker for heterochromatin H3K9me3 shows segregation at domain boundaries. Thus, boundaries act to mark the end points of heterochromatic spreading (Dixon et al., 2012). Topologically associated domains are associated with chromatin modification as loci within a domain are much more correlated with a certain chromatin modification than loci at similar distance but positioned at different TAD (Rao et al., 2014; Sanyal et al., 2012).

Other genomic features also correlate with TADs such as DNA replication, lamina-associated domains and chromocenter association (Bickmore, 2013). Based on these distinct genomic and epigenetic data, TADs have been classified into five different sub-compartments. Two of the active sub-compartments (A) correlate with active chromatin modification markers (H3K36me3, H3K79me2, H3K27ac, and H3K4me1), contain highly expressed genes and are depleted at the nuclear lamina. They however differ in regards to gene length and replication times, where the A1 sub-compartment completes replication earlier and has shorter genes on average than A2. The other three sub-compartments B1, B2, B3 are more diverse (Rao et al., 2014).

1.9 Chromatin Loops

Pairs of loci that interact more frequently than neighbouring and in-between loci represent chromatin loops, where two regions far away on the linear sequence of the chromosome are found in spatial proximity. Just as TADs, chromatin loops appear to be conserved both between cell types and organisms (Rao et al., 2014). The number of chromatin loops identified in the genome varies between studies. Sanyal et al. (2012) estimate it to be over 100,000 based on 5C data from 1% of the genome whereas Jin et al. (2013) identify 1,116,312 chromatin loops using HiC, both methods are based on genome average threshold. This is much higher than the approximately 10,000 loops identified by Rao et al. (2014) as they based loop calling to relative local background. Chromatin loops are believed to have a functional role in gene regulation (de Wit & de Laat, 2012; Gibcus & Dekker, 2013; Tolhuis et al., 2002). This regulating role of chromatin loops was demonstrated recently in the B-lymphoblastoid cell line GM12878. Up to 30% of the chromatin loops identified in GM12878 contained a promoter at one end and an enhancer at the other end. On average a 6-fold increase in gene expression was seen between genes whose promoter participated in a chromatin loop with an enhancer compared to those who didn't (Rao et al., 2014). Chromatin loops can also be the source of TAD formation. Loops can be found at the domain boundaries, causing the domain to be closed, termed "loop domain" compared to the open domain that does not contain a chromatin loop at its domain boundaries (Rao et al., 2014).

Certain proteins have been shown to stabilize chromatin loops, for example CTCF, cohesin and SMC3 (Splinter et al., 2006, Hou et al., 2008 and Phillips and Corces, 2009 and Rao et al 2014). The insulator protein CTCF is one of the most important factors in chromatin architecture, termed the "master weaver of the genome" (Phillips & Corces, 2009). Taking a closer look at the motif orientation of CTCF, loci participating in CTCF mediated chromatin looping Rao et al. (2014) reported that the majority (92%) of motif pairs are convergent. Indeed the orientation of CTCF motifs has been demonstrated to be important for chromatin looping, in turn affecting transcription and TAD formation (Guo et al., 2015) . A more recent study (Tang et al., 2015) built upon this, studying chromatin loops using CTCF ChIA-PET, generating over 50,000 CTCF-mediated chromatin loops in GM12878 cells with resolution in the 100s of bps. However they found a slightly lower portion of the CTCF mediated chromatin loops to be convergent (~65%) and a large population of tandem loops (~33%). The convergent loops are speculated to be more stable as they require more energy to form and contribute to the formation of TADs. On the other hand tandem loops are more dynamic within the TADs and contribute towards transcription and regulatory functions. Genes and enhancers that participate in CTCF mediated looping (anchor genes/enhancers) are almost exclusively housekeeping genes but genes and enhancers found in-between CTCF loops (loop genes/enhancers) are more tissue specific. Also anchor genes/enhancers were connected to active loop genes through RNAPII mediated connections, suggesting that CTCF mediated loops can serve as nucleation points to co-localize loop genes for coordinated transcription, forming transcription factories at the structural base of CTCF foci.

2. Project Aims

Gene expression is heavily depended on the conformation of the genome as genetic elements loop out to interact with each other. Chromatin organization is highly dynamic, varying both during the cell cycle and between different cell types. It is therefore interesting to look at differences in chromatin interactions between cell types to study the underlying control mechanism of gene expression. The goal of this project was to explore the long range interactions of the IRF4 promoter to regulatory elements responsible for the transcriptional regulation of IRF4 with the use of the chromosome conformation capture (3C) technique. This was done in both myeloma and melanoma cells lines to reveal cell specific regulatory elements participating in the regulation of IRF4 expression.

The goal of the project was threefold:

1. To set up and validate the chromosome conformation capture technique of appropriate quality and with the recommended controls.
2. Use the chromosome conformation capture technique to determine the long range interactions of the IRF4 promoter in myeloma and melanoma cell lines expressing different quantities of IRF4 mRNA. The cell lines studied were the Roswell Park Cancer Institute - Waldenstrom macroglobulinemia 1 line (RPCI-WM1), a human Waldenstrom macroglobulinemia (WM) cell line molecularly validated as pre-clinical model of WM (Chitta et al., 2013), U266B1 a human myeloma cell line, SKmel28 and 501mel a human melanoma cell lines and HEK293T, a human embryonic kidney cells transformed with large T antigen. The HEK293T cells were used as a negative control as they should not express IRF4 (Bairoch, 2016).
3. Regions looping to the IRF4 promoter were further evaluated for enhancer activity using histone modification ChIP, IRF4 ChIP and transcription activation assays.

3. Methods

3.1 Cell Culture

The suspension cell lines RPCI-WM1 and U266 were cultured in Roswell Park Memorial Institute medium (RPMI) supplemented with fetal bovine serum (10% and 15% final concentration, respectively) at 37°C and 5% CO_2. Both cell lines were kept at density of $5x10^5$ -$1x10^6$ cells/ml. The adherent cell lines HEK293T, SKmel28 and 501mel were cultured in Dulbecco´s Modified Eagle Medium (DMEM) supplemented with fetal bovine serum (10% final concentration) at 37°C and 5% CO_2. At about 80% cell confluency the cells where trypsinized and passaged. Cells were briefly washed with 1x PBS, filling the bottom of the flask with trypsin-EDTA and the flask incubated at 37°C for 3 minutes and media added at 3 times the volume of the trypsin-EDTA used. Cells were collected and centrifuged for 3 minutes at 1200 RPM, supernatant discarded and the cell pellet re-suspended in appropriate volume of serum containing media.

3.2 Primer and Probe Design

For 3C Gel quantification (Table 1): Primers were designed within 200 bp from *HindIII* restriction sites and the size of the PCR product kept within 300 bp. The Primer3 tool http://bioinfo.ut.ee/primer3-0.4.0/primer3/, (v.0.4.0) was used with the following parameters: primer size 28-30 bp, T_m close to 60°C (57-60°C) and 2°C maximum difference between all primers in a single experiment.

For 3C-qPCR and 3C-ddPCR quantification (Table 2): Primers were designed within 100 bp from the *HindIII* restriction site and the size of the PCR product kept within 150 bp. The Primer3 (v.0.4.0) tool was used with the following parameters: primer size 20-22 bp, T_m close to 60°C (57-60°C) and 2°C maximum difference between all primers in a single experiment. The probe on the IRF4 promoter *HindIII* fragment was designed between the restriction site and the primer on the constant fragment (qP). The Primer3 (v.0.4.0) tool was used with the following paramters: T_m 8-10°C higher than T_m of primers (68-70°C), 20-30 bp long with the following manual guidelines; more C than G residues, not containing the same base 4 times in a row and G was not allowed at the 5' most position.

Table 1. Primer sequences of HP and H1-11 used in 3C Gel quantification.

Primer/Probe	Sequence (5´-3´)
HP	GGAATGGTCCAAAGAGAACTGGAGGTG
H1	CATACTAACAGGATGGAAAGCACCCAGC
H2	CTGGGTCTTTGGGTAACTTGTTTGCAG
H3	TGAGGAATAGATACCATGGGCGTAGACAC
H4	ATGCCTTTTCCATGTCTGATTCCCTGG
H5	CCACAGTGACCCCAGTATTTTCACATCAC
H6	TTCTCTCTCTATGGGAGCAGAGCATGT
H7	CTTTTAGAGAGGAGAATTTGGGCTGGG
H8	TTTACTTGAACACTCATGGGTGGCTGC
H9	AGGTTCAGGAAAGCATAAGTCGCTCTG
H10	GCAACAACCACACTAAGAACAAGGGTC
H11	TTTCCAAAAAGGCTGCCCATCAGTGTC

Table 2. Primer sequences of qProbe, qP and q1-19 used in 3C-qPCR and 3C-ddPCR quantification.

Primer/Probe	Sequence (5´-3´)
qProbe	FAM-AGCTTCGTCATATGGCTAAACCTGGCA-BHQ
qP	GGTGGGTAAAAGAAGGCAAA
q1	GCAAGAAAGCAAATGCATGA
q2	TGAAACATCTGACTCTGCTGAC
q3	CCTACCCTCTTCCTGCTGAC
q4	TGATTTGGCACATTTTCTCATT
q5	TTCAGACTCTGCAAACCACAA
q6	TTTGTGGCTAGAATGTGGTCA
q7	TTTTGGGAACAAAAGGAAGG
q8	TCTTCTGGTGAGTTCAGTTTCG
q9	AGAACTGCCAGCAGGTAGGA
q10	CAGCGACCTCACTGCTATTG
q11	TCCATGTAAGGATTGCACGTT
q12	GCAGGCTAAGATCCCATGAA
q13	GCACTTCCTGTCTGTGCTGA
q14	TGCAAACTGTTGGTCATAAAGA
q15	CCTGCAAGAATTGGGAAGAA
q16	GTTTCCATTACCTGGTGGATT
q17	ATCATTGCAGTAGCGCCTTT
q18	GCACCTGCAGGAAACAAAAC
q19	CCAACACTTACACTGCCTCTGA

For ERCC3 the primers and probe described in the paper (Markova et al., 2011)et were used (Table 3.)

Table 3. Primer sequences used in normalization of 3C data

Primer/Probe	Sequence (5´-3´)
ERCC3_Hind1	CCAGTTGTTAGGTTGGGAAAG
ERCC3_Hind2	ACAGAAGCGGTGAGGTGAGTT
ERCC3_HindProbe	FAM-CAGTTGGGTGGGCTACACAGCAGTC-BHQ

The Primer3 (v.0.4.0) tool was used to design the primers used in ChIP-qPCR; primer size 20-22bp, T_m close to 60°C (57-60°C) and product size around 80-100 bp.

Table 4. Primer sequences used in Histone markers and IRF4 ChIP-qPCR

Primer	Sequence (5´-3´)
ChIP_IRF4p_fw	CCACCTCCAGTTCTCTTTGG
ChIP_IRF4p_rev	GCGTGTAGTAGCGGGAATCT
ChIP_IRF4i4_fw	AGGGCAGCTGATCTCTTCAG
ChIP_IRF4i4_rev	ACAGGGGAATTTGCCTTCTT
CHIP_IRF4i2_fw	GGCAGTGATAGGGTCCAAGA
CHIP_IRF4i2_rev	CACCGTTGCTCAGAACGTAA
ChIP_IRF4i5_fw	TGAGCCTCAGTCTCCCTGTT
ChIP_IRF4i5_rev	GCAAAGATGAGCTGGAGACC
ChIP_HFrag4_fw	AGCTGCCTCGTTAACTTGCT
ChIP_HFrag4_rev	CCTGGTTTAGTGGGTAAGAGGA
ChIP_HFrag6_fw	GCAAGCAGCAGCAGTTTTATT
ChIP_HFrag6_rev	CCACGATGACATTTTGAATGA
ChIP_HFrag7_fw	GTGGTAGCTGTGTGGGGATAG
ChIP_HFrag7_rev	TGAAGGCTGGAGATTGATTACA
ChIP_HFrag13_fw	CCTGGGTTTTGTGTTAAGCTG
ChIP_HFrag13_rev	AGTCGGTTCCTCTCCTCTCAC
ChIP_HFrag15_fw	TGGTACGTTGAGTGCTGAGTG
ChIP_Hfrag15_rev	GGGACTCTAAGGACCTTGTTCA
ChIP_16qControl_Fw	GACCGCGTTCAGTCCATTTAG
ChIP_16qControl_Rev	ATCTCACAGGGCAGATAGGG
ChIP_q16E1_fw	TGTTCATCGAAACAGCTTGC
ChIP_q16E1_rev	TACCCACAGTCCTGGTCCTT
ChIP_q16E2_fw	ATCCTGTCCCCACTCAGATG
ChIP_q16E2_rev	AGCACGTCAGAACACACTGC
ChIP_TFEB_fw	GCAGTGTCTGGCAGCTTCT
ChIP_TFEB_rev	GCTGTCTAGTGGCCTCTGC
ChIP_actin_fw	AGTGTGGTCCTGCGACTTCTAAG
ChIP_actin_rev	CCTGGGCTTGAGAGGTAGAGTGT

3.3 Chromosome Conformation Capture

3.3.1 3C Sample Generation

To determine the spatial organization of the IRF4 locus the protocol developed by Hagége et al. (2007) was used with minor adaptations.

Cell preparation: For adherent cell cultures (HEK293T, SKmel28 and 501mel) the medium was removed and cells washed twice with PBS. The cells were trypsinized by adding 4 ml of 0.05% (w/v) trypsin per 75cm^2 flask and incubating for 5 min at 37 °C. Cells were collected by spinning for 3min at 1200rpm at RT and then resuspended in 10% (v/v) FBS/DMEM. The cells were counted and diluted in appropriate volume of 10% (v/v) FBS/DMEM to get get a final density of 1x10^6 cells/ml. For suspension cell lines (U266 and RPCI-WM1) the cells were collected by centrifugation for 3min at 400*g* and resuspended in RPMI media containing serum. The cells were counted and medium added to reach 1x10^6 cells/ml. 10ml of the cell suspension were centrifuged for 1 min at 400g at room temperature. The supernatant was discarded and the pellet resuspened in 500 ml of 10% (v/v) FBS/PBS per 1x10^7 cells and for the SKmel28 cell line; filtered through 30 um Pre-Seperation Filter (#130-101-812 from Miltenyi Biotec) to make a single-cell suspension.

Chromatin cross-linking: For 1x10^7 cells, 9.5 ml of 1% formaldehyde/10% FBS/PBS solution were added to the single-cell suspensions and used to fix the cells for 10 min at RT. The reaction tubes were transferred to ice and 1.43 ml of 1M glycine (ice cold) added. Cells were harvested with 8 min spin at 225g at 4°C and resuspened in 5 ml cold lysis buffer (10mM Tris-HCl, pH 7.5; 10mM NaCl; 5mM MgCl2; 0.1mM EGTA; 1mM PMSF; 1x Protease Inhibitor Cocktail (#P8340 from Sigma Aldrich)) and incubated for 10 min on ice. Nuclei were harvested by centrifuging for 5 min at 400g at 4 °C and the supernatant removed. The pelleted nuclei were frozen in liquid nitrogen and stored at –80°C for few days or weeks at this stage.

Digestion: The nuclei were resuspended in 0.5 ml of 1.2x restriction enzyme buffer (NEBuffer 2.1 (#B7202S) for *HindIII*) and SDS added to final concentration of 0.3% and incubated for 1 h at 37 °C while shaking at 900 rpm. Triton X-100 was added to a final concentration of 2% and further incubated for 1 h at 37°C while shaking at 900 rpm. A 25µl aliquot (for gel digestion efficiency estimation) of the sample was taken as undigested genomic DNA control. Then, 400U of *HindIII* (#R0104 from NEB) were added to the sample and incubated overnight at 37 °C while shaking at 900 rpm. Again a 25µl aliquot of the sample were taken as a digested genomic DNA control. If 25µl aliquots were taken, SDS and Triton-X-100 were added to keep the final values at 0.3% SDS and 2% Triton-X-100.

Ligation: SDS was added to the sample to final concentration of 1.6% and incubated for 20-25 min at 65°C while shaking at 900 rpm The digested nuclei were transferred to a 50 ml falcon tube containing 6.12ml of 1.15x T4 DNA ligation buffer (#B0202S from NEB) and triton X-100 to a final concentration 1% and the sample incubated for 1 h at 37°C while shaking gently. Then 3.35 µl T4 ligase (6700 CELU) (#M0202M from NEB) was added and again incubated for 4 h at 16 °C followed by 30 min at room temperature (18–22 °C). Then 300 µg of Proteinase K (#EO0491 from Thermo Scientific) were added and the sample incubated overnight at 65°C to reverse-crosslink the sample.

DNA purification: To start the DNA isolation, 300µg of RNase A (#EN0531 Thermo Scientific) were added to the reverse crosslinked sample and incubated for 30-45 min at 37°C. Then 7ml of phenol–

chloroform were added, mixed vigorously, and centrifuged for 15 min at 2,200g. The supernatant was transferred into a new 50 ml tube and 7 ml of distilled water, 1.5 ml of 2 M sodium acetate pH 5.6 were added followed by 35 ml of ethanol. This mixture was mixed and placed at -80°C for approximately 1 h, centrifuged for 45 min at 2,200g at 4 °C, the supernatant removed and 10 ml of 70% (v/v) ethanol added. The mixture was centrifuged for 15 min at 2,200g at 4 °C, the supernatant removed and the pellet briefly dried at RT before dissolving the DNA pellet in 150 µl of 10 mM Tris pH 7.5.

Second digestion step: 1µl sample was taken for ligation quality check and 30µl of ddH_2O, 20µl of 10x 1.1NEBuffer for the selected restriction enzyme and 1µl (or 100U) of *SacI* added. this was incubated for 2 h at 37 °C and then column purified using PCR clean up kit (#740609 from MACHEREY-NAGEL). The DNA concentration was measured with Nanodrop and the sample diluted in 10 mM Tris pH 7.5 to get a final concentration of 100 ng/µl.

3.3.2 3C Digestion Efficiency Analysis

In order to analyze the efficiency of the 3C digestion, 500 µl of 1x PK buffer (5 mM EDTA, pH 8.0; 10 mM Tris-HCl, pH 8.0; 0.5% SDS) and 60 µg of Proteinase K were added to the 25 µl digested and undigested samples taken earlier and incubated for 30 min at 65°C to reverse the crosslinks. Then 5µg of RNase A were added to the samples and incubated for 2h at 37°C. Equal volumes of phenol–chloroform were added, centrifuged for 5 min at 16,100g and the supernatant transferred to a new tube. The DNA was precipitated by adding 1/10 volume of 2 M sodium acetate pH 5.6, glycogen (10µg) and 2x volume of pure ethanol to the samples and incubating at -80°C for ca 45 min. The DNA pellet was collected and washed with 70% ethanol and resuspended in 20µl of ddH2O. The Dna was analyzed by running the samples on an 0.6% agarose gel using 10-20µl of the purified undigested and digested samples.

3.3.3 Ligation Efficiency

Proper ligation and additional restriction enzyme digestion: To see if the ligation and the secondary digestion step were successful,1 µl sample was taken before the additional restriction enzyme digestion step and about 500ng of the final 3C sample were run on a 0.6% gel along with DNA ladder, high range (#SM1351 from Thermo Scientific).

3.3.4 Assessment of sample purity

Aliquots of the 3C sample were diluted two- to fourfold in a serial dilution. Genomic DNA was added to the diluted 3C samples such that the total amount of DNA in each reaction was constant (400 ng). A probe-based qPCR reaction (20µl) (JumpStart Taq ReadyMix for qPCR from Sigma, #7440) was set up as detailed in Table 5 with diluted 3C sample using the qP +q10 primer set.

Table 5. 20µl probe-based qPCR to assess sample purity

	Vol (µl)
2x Readymix	10
Fw Primer (10µM)	0.3
Rev Primer (10µM)	0.3
$MgCl_2$ (25µM)	2.8
Probe (5µM)	0.5
H_2O	2.1

3.3.5 Quantification of 3C samples

3C-Gel Quantification: To quantify the 3C samples, 1.5µl of each sample and 1µl of 20x diluted BAC control template were used in PCR reactions with primers HP + H1-11 following the recommended setup for 25µl PCR with Taq polymerase (#1007837 from Qiagen). After the PCR, the products were run on 2% agarose gels for 40 minutes at 90 volts and the intensity of the products on the gel measured with Genesnap (v7.02.01) from Syngene.

3C-qPCR: For each reaction, 4µl of each 3C sample (100ng/µl) were used in each 20µl probe based qPCR setup (table 5.) with a JumpStart Taq ReadyMix for qPCR on a 96-well plate, spun down for 10s at 1k rpm and measured with AB 7500 real time PCR system. For each primer pair used in the quantification the slope and the intercept were calculated from standard curves of the pair with the BAC control template. These standard curve parameters were then used to calculate the interaction frequency between the two fragments that these primers correspond to, using the formula (a=slope, b=intercept):

$$IF = 10^{(c_t - b)/a}$$

3C-droplet digital PCR (3C-ddPCR): For each 3C sample, 4µl (100ng/µl) were used in each 22µl reaction in the following ddPCR setup with a probemix from BioRad (#1863010 from BioRad). A similar setup was also done for ERCC3 using the ERCC3 primers 1 and 2 and the ERCC3 probe.

Table 6. 22µl ddPCR setup for 3C quantification

	Vol (µl)
2x ProbeMix	11
qP (10µM)	0.6
q1-19 (10µM)	0.6
$MgCl_2$(25µM)	0.5
qProbe (5µM)	0.5
H_2O	3.8

The 22µl ddPCR solutions were mixed well by pipetting 15 times up and down. The 96-well plate was spun down for 30 seconds at 1000 rpm and placed in the BioRad Automatic Droplet Generator to divide each sample in a 96-well PCR plate into ~20,000 uniform nanoliter droplets. The plate was sealed and the recommended PCR program run to the end point in a PCR machine. The droplets in each

sample were then analysed using a QX200 droplet Reader from BioRad which measures in each droplet either a fluorescence signal or not. This data was then analysed (Quantasoft, v1.7.4) using poissons statistics to determine the target DNA template concentration in the original sample. Only one technical replicate was done using ddPCR and the data was not corrected for primer efficiency. Instead ddPCR gives an absolute number of ligation products in each reaction which was directly normalized to the interaction frequency control ERCC3 to get interaction frequency value of each product.

3.3.6 Creating the BAC Control Template

BAC DNA isolation: To isolate DNA from the BAC CH17-440M10 (IRF4-BAC) clone, a 250ml culture of transformed *E.coli* was DNA purified using scaled down version of the alkaline lysis method, followed by double acetate precipitate protocol (Roe., 1999).

To make the ligated BAC control template the protocol from Naumova et al. (2012) was followed.

Digestion**:** The purified BAC DNA was digested with 4µl *HindIII* in a 500µl reaction overnight. Phenol-chloroform extraction was completed, followed by ethanol precipitation and the pellet of digested BAC DNA was briefly air dried and resuspended in 44µl of ddH_2O.

Ligation: The digested BAC DNA (43µl) was ligated with T4 DNA ligase in a 60µl reaction at 16°C overnight. After ligation, the reaction was incubated at 65°C for 15 minutes. The volume was increased to 200µl with ddH_2O and phenol-chloroform extraction done again followed by ethanol precipitation. The ligated DNA was briefly air dried and resuspended in 100µl of TE, pH 8.0.

Standard Curve setup: The BAC control template was diluted in four 10-fold serial dilutions. Probe based qPCR was done using Sigma JumpStart Taq ReadyMix for qPCR following the setup in table 5 plus 5µl of diluted BAC control sample, done in triplicates. The slope and the intercept were calculated for each 3C primer pair (q1-17) from the Ct values plotted against the logarithm of the amount of BAC control used. Primer efficiency was then calculated by:

$$E = (10^{\frac{-1}{slope}} - 1) \cdot 100\%$$

3.4 RNA isolation and RT-PCR

Total RNA was extracted from $1x10^6$ cells (RPCI-WM1, U266B1, SKmel28, 501mel and HEK293T) in biological triplicates with 1ml of TRI reagent solution (#AM9738 from Ambion) followed by chloroform extraction and isopropanol precipitation. The resulting RNA was dissolved in 20µl nuclease-free water and RNA concentration determined using Nanodrop. 1µg of total RNA was reverse transcribed using the GoScript™ reverse transcriptase system (#A5001C from Promega).

Gene expression was analysed using qPCR. For each qPCR reaction, 50ng of cDNA, 0.3µM of primers (Table 7) and the appropriate volume of SYBR® Green JumpStart™ Taq ReadyMix™ (#S4438 from Sigma Aldrich) was added to the reaction according to the manufacturer's instructions. IRF4 expression was normalized to actin and expression of IRF4 in each cell line was calculated as a fold change of the IRF4 expression in HEK293T cells. Statistical significance of the difference in IRF4 expression between RPCI-WM1 and U266B1 was tested using unpaired t-test in R (v. 3.2.0).

Table 7. IRF4 and Actin expression primers used in the RT-PCR protocol.

Primer	Sequence (5´-3´)
IRF4_exp_fw	ATGTCCATGAGCCACCCCTA
IRF4_exp_rev	TAGTTGTGAACCTGCTGGGC
Actin_exp_fw	AGGCACCAGGGCGTGAT
Actin_exp_rev	GCCCACATAGGAATCCTTCTGAC

3.5 Histone modification chromatin immunoprecipitation (HM-ChIP)

Based on the Young lab protocol for ChIP (Lee et al., 2006).

For crosslinking, $4x10^6$ RPCI-WM1 cells were collected in 4ml of RPMI media and 100µl of 16% formaldehyde (final concentration 0.4%) added and left at RT on shaker for exactly 10 min. Then 200µl of 2.5M glycine were added and left on a shaker for 5 min. Crosslinked cells were spun down at 4**°**C at 1500rpm for 5 min and washed twice with PBS and then centrifuged at 8000rpm for 30 sec and the pellet snap frozen with liquid nitrogen and kept at -80**°**C.

For antibody coating, 50µl of magnetic beads (Dynabeads #10006D from Thermo Scientific) were washed 3x with 1ml of dilution buffer (0.01% SDS, 1.1% TX-100, 1.2mM EDTA, 16.7mM Tris pH 8.0, 167mM NaCl, with Proteinase inhibitors Cocktail, 1mM PMSF and 50µl 1M Na-Butyrate) on a magnetic stand. The washed beads were divided into four tubes with 170µl each and 5ug of one of the following antibodies from Cell Signaling added PAN/H3 (#4620), H3K4Me3 (#9751), H3K4Me1 (#5326) and H3K27Ac (#8173) and the mixture rotated at 4°C for 1h and washed once with dilution buffer before the remaining buffer was removed.

For sonication, $1x10^6$ of crosslinked cells were resuspended in 1ml lysis buffer (10 mM Tris, 10mM NaCl, 0.2% NP40 with 10µl/ml Proteinase Inhibitors and 1mM PMSF and 5mM Na-butyrate) and left on ice for 5 minutes. The lysate was centrifuged at 2000rpm for 5 min at 4°C and the supernatant removed. 100µl of dilution buffer were added to the pellet and sonicated 3x for 5 min, on high with 30 seconds on and 30 seconds off, using the Bioruptor from Diagenode. 400µl of dilution buffer and washed beads were added to each tube of sonicated materials, and rotated in cold room for 1 hour to pre-clean the samples. The pre-cleaned samples where then magnetized, the beads removed and all four lysates pooled together. Then 320µl were taken as input and the rest was divided between the four tubes of antibody-coated beads and rotated overnight at 4°C.

Washing: Working at 4°C the samples were magnetised, the supernatant removed and washed once with low salt wash buffer (0.1% SDS, 1% TX-100, 2mM EDTA, 20mM Tris pH8.0, 150mM NaCl), once with high salt wash buffer (0.1% SDS, 1% TX-100, 2mM EDTA, 20mM Tris pH 8.0, 500mM NaCl) and once with LiCl wash buffer (250mM LiCl, 1% NP40, 1% Na-DOC, 1mM EDTA, 10mM Tris pH8.0) after 5 min rotation with each wash buffer. Washed beads were resuspended in TE buffer and samples transferred to new tubes and eluted in 150µl elution buffer for 10 min at 68°C. Proteinase K (150 µg) was added to each sample and to the input. The samples and the input were reverse crosslinked at 42°C for 2h then 68°C for 6h and kept at 4°C overnight.

Purification: The reversed crosslinked samples and input DNA were purified with phenol-chloroform extraction and ethanol precipitation and dissolved in 40µl of nuclease-free water.

Quantification: The ChIP samples were quantified with qPCR using SYBR® Green JumpStart™ Taq ReadyMix™ (#S4438 from Sigma Aldrich) and 0.3µM of the primers in Table 4 according to the manufacturer's instructions.

3.6 IRF4 chromatin immunoprecipitation

Based on the Young lab protocol for ChIP (Lee et al., 2006).

For crosslinking, $4x10^7$ RPCI-WM1 cells were collected in RPMI media and 16% formaldehyde (final concentration 0.4%) added and left at RT for exactly 10 min. Then 2.5M glycine equal to the 1/20 of the crosslinking volume was added and left on a shaker for 5 min. Crosslinked cells were spun down at 4°C, 2500rpm for 10 min and washed twice with ice cold PBS and then centrifuged at 4000rpm for 10 min. The pellet was then snap frozen with liquid nitrogen and kept at -80°C.

In order to preblock and and allow antibody binding to beads, 50µl of magnetic beads (Dynabeads) were washed 2x with 1ml of 0.5% BSA/PBS solution on a magnetic stand. The washed beads were divided into two tubes with 250µl of 0.5% BSA/PBS solution and 15µg of either IgG (#sc-2028 from Santa Cruz) or IRF4 (#sc-609 from Santa Cruz) antibody, rotated at 4°C for overnight and washed 3x with 1ml of 0.5% BSA/PBS solution before the remaining buffer was removed.

In order to sonicate the DNA, $4x10^7$ of crosslinked cells were resuspended in 5 ml of lysis buffer 1 (50 mM Hepes-KOH, pH 7.5, 140 mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40,0.25% Triton X-100 with 10µl/ml Proteinase Inhibitors and 1mM PMSF) and rocked at 4°C for 10 minutes. The lysate was centrifuged at 4000rpm for 5 min at 4°C and the supernatant removed. 5ml of lysis buffer 2 (200 mM NaCl, 1mM EDTA, 0.5mM EGTA, 10mM Tris pH 8 with 10µl/ml Proteinase Inhibitors and 1mM PMSF was added and the lysate rocked at 4°C for 5 minutes and then centrifuged at 2500 rpm for 5min at 4°C. Finally, 900µl of lysate lysis buffer 3 (1mM EDTA, 0.5mM EGTA, 10mM Tris-HCl pH8, 100mM NaCl, 0.1% Na-Deoxycholate, 0.5% N-lauroyl sarcosine with 10µl/ml Proteinase Inhibitors and 1mM PMSF) were added and 300µl of the lysate used in each sonication reaction which involved the following steps: 5min 15s on/30s off at 25% output using a sonicator from Active Motif (model Q120AM). The 300µl sonicated samples were pooled and lysis buffer 3 added to 1ml and 100µl of 10% Triton X-100 added. The debris was spun at 14K for 10min. At least 1/50 of the total volume was saved as input control and stored at -20°C. The rest of the sonicated lysate was partitioned into two tubes and antibody prebound beads (IRF4 and IgG) were added into each tube and the mixture rotated overnight at 4°C.

Washing: Working at 4°C the samples were magnetised, the supernatant removed and the beads washed twice with low salt wash buffer (0.1% SDS, 1% TX-100, 2mM EDTA, 20mM Tris pH8.0, 150mM NaCl), once with high salt wash buffer (0.1% SDS, 1% TX-100, 2mM EDTA, 20mM Tris pH 8.0, 500mM NaCl), once with LiCl wash buffer (250mM LiCl, 1% NP40, 1% Na-DOC, 1mM EDTA, 10mM Tris pH8.0) and once with TE plus 50mM NaCl, after 5 min rotation with each wash buffer. The beads were centrifuged at 3k for 3 min and any residual buffer was aspirated. 100µl of elution buffer were added and the DNA-protein complexes eluted from the beads at 65°C for 15min with 1000rpm shaking.

Beads were removed by centrifuging at 14k for 1 min and the supernatant reverse crosslinked overnight at 65°C along with the input control.

Purification: After RNase A (0.2µg/µl at 37° for 2h) and proteinase K (0.2µg/µl, at 55°C 2h) treatments the reversed crosslinked samples and input DNA were purified with phenol-chloroform extraction and ethanol precipitation and dissolved in nuclease-free water.

Quantification: The ChIP samples were quantified with qPCR using SYBR® Green JumpStart™ Taq ReadyMix™ (#S4438 from Sigma Aldrich) and 0.3µM of the primers ChIP_16qControl, ChIP_q16E1, ChIP_q16E2, ChIP_TFEB and ChIP_actin in Table 4 according to the manufacturer's instructions.

3.7 Transcription activation Assay

Construction of pGL3-promoter-16qE1 and pGL3-promoter-16qControl vectors: Two fragments on chromosome 6, 16qE1 (chr6:341143-341545) and 16qControl (chr6:344263-344668) were amplified using Q5 High-Fidelity DNA Polymerase (#M0491S from NEB) following the recommended protocol using the primer pairs in table 6.The primers used to clone the 16qE1 and 16qControl fragments into the pGL3-promoter were designed using the NEBuilder assembly tool (v1.11.2), minimal primer length of 18 bp and minimal overlapping of 15 bp with the pGL3-promoter vector (#E1761 from Promega).

Table 8. Primer sequences used in Gibson assembly cloning of 16qE1 and 16qControl regions into pGL3-promoter vector.

Primer	Sequence (5´-3´)
pGL3p_16qE1_fw	cgataggtaccgagctTAGCCCCGGGCGCTGTTT
pGL3p_16qE1_rev	gctagcacgcgtaagATGCCTTTTCCAAGAAAAGCCC
pGL3p_16qControl_fw	cgataggtaccgagctGTCATCATTTAGAACTTAGTTCTATTATTTG
pGL3p_16qControl_rev	gctagcacgcgtaagATAAAGTGGGGCTGGTATC

Gel electrophoresis was used to confirm the appropriate size of the amplified fragments and the fragment then purified using PCR clean-up kit from MACHEREY-NAGEL. The resulting PCR fragments were cloned into the pGL3-promoter luciferase reporter vector (Promega) with Gibson assembly (#E2611 from NEB) following the recommended protocol from NEB. The correct sequence of each of the inserts in the newly made plasmids, pGL3-promoter-16qE1 and pGL3-promoter-16qControl was confirmed with Sanger sequencing performed by Beckman Coulter Genomics using the universal primer RVprimer3 (5'-CTAGCAAAATAGGCTGTCCC-3').

Transfection of RPCI-WM1 cells: RPCI-WM1 cells were washed and resuspended in RPMI 1640 containing 10% FBS. About $1x10^6$ cells (90µl) were placed in a 4mm electroporation cuvette with 9µg of one of the pGL3 vectors (table 6) along with 1µg of pCMV Renilla and electroporated using a Gene Pulser Xcell (BioRad) with the parameters 350µF and 220V. After electroporation the cells were immediately added to 900µl of the appropriate medium and incubated at 37°C and 5% CO_2. The cells were lysed after 48h and the firefly luciferase and renilla luminescence activity was measured using the Dual-Glo kit (#E2940 from Promega).

Transfection of HEK293T and 501mel cells: HEK293T and 501mel cells were washed and resuspended in DMEM containing 10% FBS. $2x10^4$ cells were seeded per well on a 96-well plate and incubated at 37°C and 5% CO_2 for 24h. The cells were transfected with 90ng of one of the pGL3 vector (table 6) and 10ng of pCMV-renilla using FuGENE (#E2311 from Promega) following the recommended protocol from Promega. The cells were lysed for 48h and the firefly luciferase and renilla luminescence activity measured using the Dual-Glo kit from Promega.

Table 9. pGL3 vectors used in the luciferase assay

pGL3 vectors
pGL3-Control
pGL3-Promoter
pGL3-Promoter-16qE1
pGL3-Promoter-16qControl
pGL3-Promoter-Fragment 4

4. Results

4.1 Chromosome Conformation Capture (3C) Setup

4.1.1 Quality Controls

Following the protocol from Hagége et al. (2007), the 3C method was setup successfully with minor modifications as described in the Methods section (3.3). Three different quality controls were used to confirm successful 3C experiment for each cell line used in the protocol. Each 3C sample that passed the quality controls was quantified.

4.1.1.1 Digestion Efficiency

The first crucial checkpoint in the 3C protocol is confirming an effective digestion of the 3C sample. Ineffective digestion of 3C samples leads to large DNA fragments and inaccurate quantification of interacting chromatins. Digestion efficiency was estimated by gel electrophoresis for all cell lines used. It is important to choose a restriction enzyme that results in an even fragment size distribution (1-10kb) around regions of interest which here is the IRF4 locus and especially the IRF4 promoter. Digestion by the restriction enzyme *SacI* was tested originally as it gave the most favorable fragment pattern around the IRF4 promoter. The optimization of *SacI* digestion was however unsuccessful (Figure 1) as different digestion conditions failed to result in a complete digestion of the fixed sample. It has been suggested that enzymes that require low salt buffers, like *SacI*, are not compatible with the 3C protocol (Gavrilov et al., 2009). The second best option, *HindIII* digestion worked well for all cell lines tested (Figure 1).

4.1.1.2 Ligation Efficiency

Another important quality check of the 3C protocol is to check if the ligation and the secondary digestion step were successful before starting the quantification of ligation fragments. An excellent 3C template should run as one tight band above 10kb in size on the agarose gel. A mediocre but acceptable 3C template can look like a tight band with a noticeable smear running down the gel. Smeary templates can be successfully used in 3C experiments but completely degraded templates should be excluded from further analysis (Naumova et al., 2012). Examples of the ligated 3C samples for the cell lines used in this study can be seen in Figure 2.

4.1.1.3 Purity assessment

The final quality control is a sample purity assessment which was done by diluting the 3C samples in two-fold serial dilutions. The diluted 3C samples were then quantified using qPCR as described in the Methods section 3.3.4. Examples of a 3C sample with acceptable purity and a 3C sample with unacceptable purity are shown in Table 10 for two 3C biological replicates of 501mel.

While there is no absolute cut off factor, samples that deviate from expected dilution pattern, such as the 501mel dBR4 sample, were discarded and samples that followed the expected dilution factor were kept (501mel dBR5).

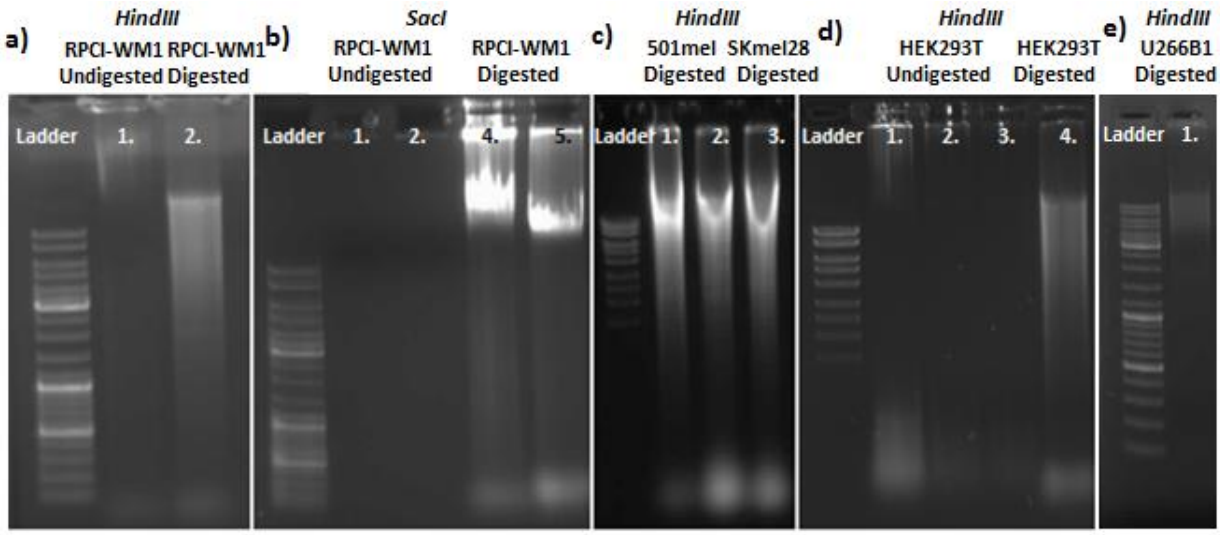

Figure 1. HindIII Digestion Efficiency of 3C samples.

HindIII digestion of 3C samples on 1% agarose gel from the cell lines RPCI-WM1, 501mel, SKmel28, U266B1 and HEK293T compared to DNA ladder (10kb-1.5kb). a) Comparison between undigested (lane 1) and *HindIII* digested (lane 2) 3C samples of RPCI-WM1 demonstrating successful digestion. b) Comparison between undigested (lanes 1 and 2) and *SacI* digested (lanes 3 and 4) 3C samples of RPCI-WM1 demonstrating unsuccessful digestion. c) *HindIII* digested 501mel (lanes 1 and 2) and SKmel28 (lane 3) 3C samples demonstrating modest digestion. d) Comparison between undigested (lanes 1, 2 and 3) and *HindIII* digested (lane 4) 3C sample of HEK293T demonstrating successful digestion. e) *HindIII* digested U266B1 (lane 1) 3C samples demonstrating successful digestion.

Table 10. Sample purity assessment for two 3C biological replicates of 501mel.

Sample Name		501mel dBR4			501mel dBR5		
Dilution factor		1x	2x	4x	1x	2x	4x
b (intercept)		36.38	36.38	36.38	36.38	36.38	36.38
a (slope)		-3.33	-3.33	-3.33	-3.33	-3.33	-3.33
Ct		31.53	32.89	34.39	31.54	32.40	33.41
Values		28.53	11.14	3.95	28.35	15.68	7.77
Dilution factor	Expected		2	4		2	4
	Observed		2.56	7.22		1.81	3.65

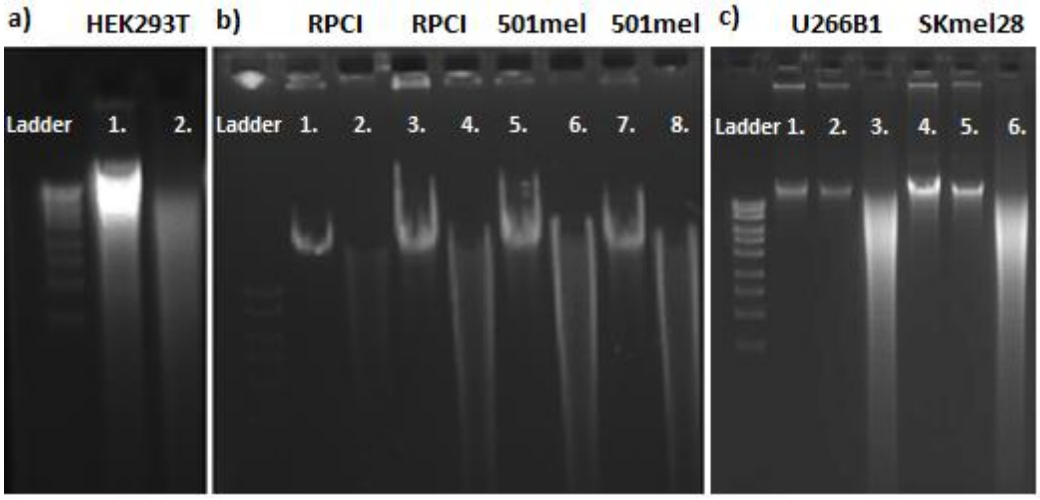

Figure 2. Ligation Efficiency of 3C samples

Examples of ligated 3C samples and secondary digestion with *SacI* of the cell lines RPCI-WM1, 501mel, SKmel28, U266B1 and HEK293T compared to DNA ladder (10kb-1.5kb). a) Example of ligation (lane 1) and secondary *SacI* digestion (lane 2) of 3C sample of HEK293T. b) Example of ligation (lanes 1, 3, 5 and 7) and secondary *SacI* digestion (lanes 2, 4, 6 and 8) of 3C samples of RPCI-WM1 and 501mel. c) Example of ligation (lanes 1, 2, 4, 5) and secondary *SacI* digestion (lanes 3 and 6) of 3C samples of U266B1 and SKmel28.

4.1.2 3C Controls

To carry out a proper analysis of 3C experiments, three controls are recommended by Dekker (2006): PCR efficiency control, assessment of background collisions and data normalization, along with two negative controls were performed. The negative controls were done in parallel with the preparation of normal 3C samples following the 3C protocol, as indicated, without using either ligase (non-ligation control) or formaldehyde (non-formaldehyde control). The non-ligation control showed no ligation products in a qPCR measurement and the non-formaldehyde control showed no products or very few products for fragments close to the promoter (q9 and q10). By designing primers H1-11 for 3C-gel quantification (Table 1) and q1-q19 for 3C-qPCR and ddPCR (Table 2) evenly around the IRF4 locus, background collisions can be accounted for.

The bacterial artificial chromosome CH17-440M10 (IRF4 BAC) was used to make a template of the *HindIII* ligation fragments (Method section 3.3.6) as it covers a 85kb region upstream and 160kb region downstream of the IRF4 promoter. PCR efficiency for primer sets qP + q1-17 (Table 2) were calculated from a standard curve of diluted BAC control template. The PCR efficiency was not calculated for primers q18 and q19 as the primers fall outside the region covered by IRF4-BAC. These primers were still used in quantification with ddPCR as ddPCR enables absolute quantification and does not require standard curves. PCR efficiency was between 90-101% for all primer pairs tested.

Another control is needed to normalize the 3C data between the cell lines as the overall efficiency of the 3C experiment might differ between the cell lines. For this a ligation fragment from the *HindIII* digested ERCC3 locus previously reported for data normalization was used (Markova et al., 2011). The PCR efficiency for the ERCC3 primer pair was calculated from a standard curve using serial dilutions of a plasmid containing a part of the same ligation fragment that the ERCC3 primer pair (Table 3) quantifies

in the 3C samples (Methods section 3.3.6). ERCC3 is a housekeeping gene and should have similar chromosome fibre interactions in all five cell lines (Hagege et al., 2007; Markova et al., 2011).

4.2 Quantification of 3C Samples with Gel Quantification

Gel quantification is an easy and accessible method to quantify and confirm the correct size of the ligation fragments. Therefore two biological replicates of cultured RPCI-WM1 cells were assayed using 3C-Gel Quantification (Method Sections 3.3.1 and 3.3.5) without column purification and second digestion in an initial attempt at setting up 3C.

11 primer pairs (HP + H1-H11) (Table 1) were designed close to *HindIII* restriction sites to enable quantification of ligation fragments around the IRF4 locus and measure the chromatin looping from the IRF4 promoter. *HindIII* digested and ligated BAC template was used as a control to normalize for differences in primer efficiency. The samples could not be accurately quantified with Nanodrop and thus the concentration used in each reaction was counted as the percent of the 3C sample. For each reaction, 1/100 of 3C samples and 1/2000 of the BAC control template were used.

The results of the gel run (Methods 3.3.5) can be seen in Figure 3A. The values for the BAC control template products were used to normalize for differences in PCR efficiency between primer pairs giving interaction frequency values which can be plotted against primer location on chromosome 6 (Figure 3B). Most ligation products can be seen as singular bands, although extra bands can be seen for H1, H4 and H10 (Figure 3A). In addition primer dimer formation was frequently observed. The efficiency of the PCR was poor as 45 cycles were needed to get visible bands and even then some bands were faint (H11 and H1).

The chromosome undergoes fluctuations causing random chromosome fibre interactions between fragments on the chromosome. These random interactions decrease with increased distance on the linear scale. As expected, chromosome fibre interaction frequencies to the IRF4 promoter decrease with distance from the promoter. Two peaks of high chromosome fibre interactios with the IRF4 promoter can be located 31 kb upstream and 95 kb downstream of the IRF4 promoter which might be indicative of functional chromatin loops between those regions and the IRF4 promoter. The interaction frequency for H1 should be interpreted with caution as the primer pair is most likely defective as a faint band also appears with the BAC control template. The inconsistency between the two technical replicates is apparent and is indicative of a problem with the gel quantification, extra bands and/or impure 3C samples. This makes it difficult to draw conclusions from this experiment.

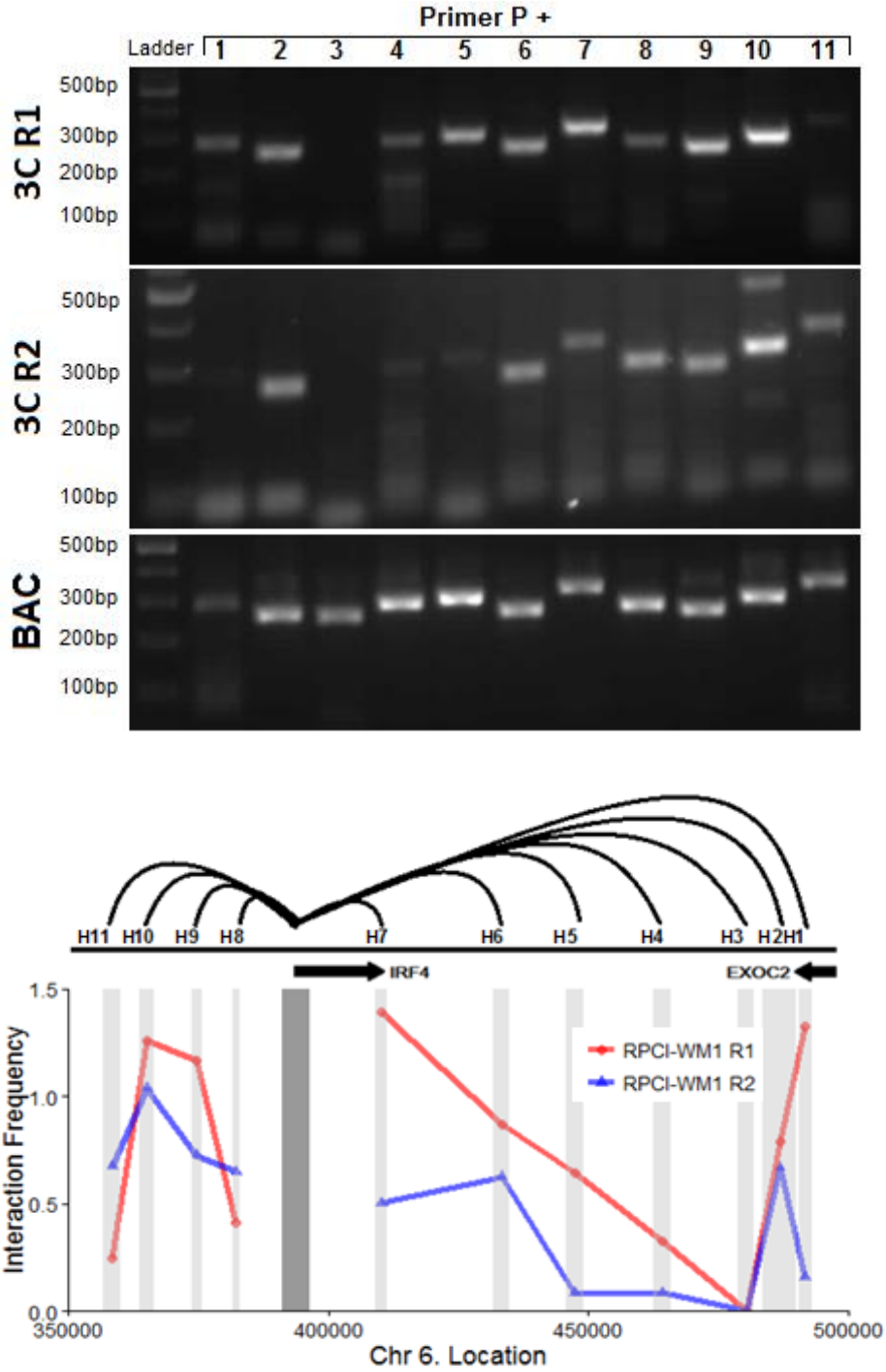

Figure 3. 3C-Gel Quantification

a) Gel run of PCR products obtained with primers HP + H1-11 on agarose gel using the digested and ligated IRF4 containing BAC (CH17-440M10) and 3C samples of RPCI-WM1 as template for the PCR in two biological replicates (R1 and R2). b) The intensity values of the ligation products from the BAC were used to normalize the intensity values of the ligation products from two technical replicates (R1 and R2) of a 3C sample of RPCI-WM1. Relative interaction frequency was determined and plotted against the location on chromosome 6.

4.3 Quantification of 3C samples with qPCR

Inconsistencies in quantification and low PCR efficiency using 3C-gel-quantification caused us to switch to quantification with probe based qPCR. qPCR offers increased specificity and allows for a better control for differences in primer efficiencies with the use of standard curves of a diluted BAC control template. qPCR was used to quantify ligation products in two new biological replicates of the RPCI-WM1 cell line and one replicate of the HEK293T cell line using the 17 qPCR primer pairs (Table 2). For each primer pair used in the quantification, the slope and the intercept along with the primer efficiency were calculated from a standard curve of a serial diluted BAC template (Methods section 3.3.6). These standard curve parameters were then used to calculate the interaction frequency between the two fragments that these primers correspond. The results were normalized to the IRF4-promoter q9 interaction at this stage and plotted according to their location on chromosome 6 in Figure 4.

Using 3C qPCR quantification, the two biological replicates of RPCI-WM1 align better than with 3C-Gel quantification (Figure 3) although the interaction frequency for q15 is inconsistent. In the RPCI-WM1, cells a clear elevation was observed of q15, q16 and q17 interaction frequencies to the IRF4 promoter as compared to those in HEK293T, indicating a chromatin loop from this region to the IRF4 promoter in RPCI-WM1. Again, as expected, chromosome fibre interaction frequencies to the IRF4 promoter decrease with distance from the promoter in both RPCI-WM1 and HEK293T, raising confidence of successful 3C experiments. At the same time, qPCR is dependent upon successful standards which might be responsible for the aberrantly high interaction frequency at q13.

4.4 Quantification of 3C samples with droplet digital PCR (ddPCR)

Droplet digital PCR (ddPCR) provides absolute quantification of the target copies and does not rely on standards, resulting in more precise quantifications of rare target copies. Thus there is no need to consider the difference in primer efficiencies when analyzing the ddPCR data. This also enables measuring of connections from the IRF4 promoter to fragments outside of the region covered by the IRF4-BAC. 3C samples were made in biological triplicates for the cell lines RPCI-WM1, 501mel and HEK293T but only one biological replicate of U266B1 and SKmel28. 3C samples were measured using primers q1-19 (Table 2) and the ERCC3 primer pair (Table 3) using 400ng of the samples per reaction in a droplet digital PCR setup (methods section 3.3.5).

Figure 5 shows the chromatin fibre interaction frequencies for fragments q1-19 to the IRF4 promoter as normalized for ERCC3 interaction frequency for all five cell lines plotted according to their location at chromosome 6.

Again for RPCI-WM1, a clear isolated interaction peak can be seen about 40kb upstream of the IRF4 promoter. The U266B1 and the 501mel cell lines also appear to have increased chromosome fibre interactions to the IRF4 promoter at the same region. However, these interactions are less frequent than the interactions seen in RPCI-WM1. The increased chromosome fibre interactions to the IRF4 promoter in 501mel cells corresponds to the *HindIII* fragments q14, q15 and q16 however in the RPCI-WM1 and U266B1 cell lines the q15, q16 and q17 *HIndIII* fragments are elevated. As expected, the normalized interaction frequencies to the IRF4 promoter for the negative control cell line HEK293T around IRF4 account for the background generated from random collision and the interaction frequency decreases

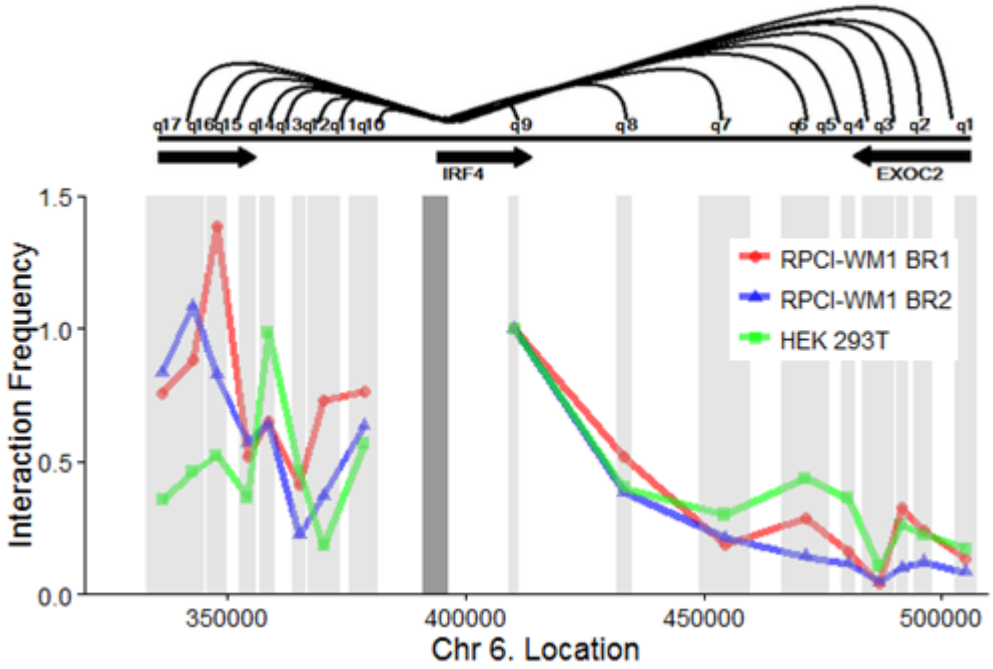

Figure 4. 3C-qPCR.

The interaction frequency values from the IRF4 promoter in two biological replicates of RPCI-WM1 (BR1 and BR2) and HEK293T cell lines to nearby *HindIII* fragments normalized to the interaction with fragment 9 plotted against the location on chromosome 6.

with increased distance from the IRF4 promoter. Downstream of IRF4 the chromosome fibre interactions to the IRF4 promoter in the cell lines RPCI-WM1, U266B1 and SKmel28 fall closely with the interactions for the negative control cell line HEK293T. However in the 501mel cell line the interaction frequency remains high for all fragments tested. This demonstrates the high chromosome fibre interaction background in 501mel cells. An exceedingly high interaction frequency was observed at q10 for the 501mel cell line and to an extent for the SKmel28 cell line. Because of how close this *HindIII* fragment is to the anchor, a clear isolated peak cannot be obtained with a 6-base cutter 3C. This makes it difficult to conclude about a potentially biologically functional loop at q10 in 501mel and SKmel28 cells.

4.5 IRF4 mRNA Expression

To put the 3C data into context it is important to quantity the expression of IRF4 in the cell lines tested with 3C to determine the effects that the chromatin fibre interactions might have on the IRF4 regulation. IRF4 expression was measured by quantifing the mRNA levels of IRF4 using RT-PCR (methods section 3.4). The B cell linage cell lines RPCI-WM1 and U266B1 contain the most IRF4 expression, 652 and 393 fold HEK293T IRF4 expression respectively (Figure 6). This indicates a strong effect of the region participating in the chromatin loop to the IRF4 promoter in RPCI-WM1 and U266B1 on IRF4 expression. However the fact that the loop is extensively stronger in RPCI-WM1 cells is not reflected on the IRF4 expression as there is not a significant difference in the IRF4 expression between RPCI-WM1 and U266B1 (unpaired t-test, $p = 0.441$).

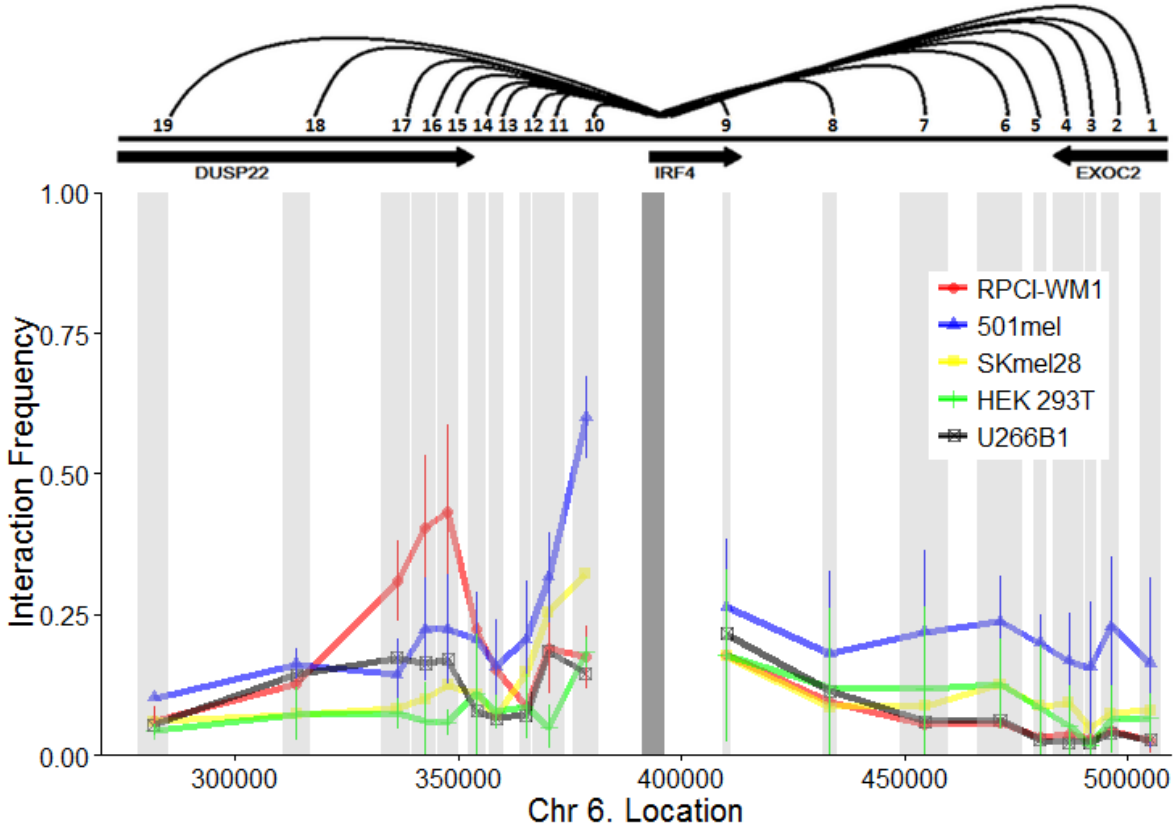

Figure 5 3C-ddPCR.

Results of chromosome conformation capture (3C) experiments performed on the IRF4 locus in RPCI-WM1, 501mel, HEK293T, Skmel28 and U266B1 cell lines shown as interaction frequency normalized by ERCC3 interaction frequency along the location on chromosome 6. Error bars correspond to standard deviation between biological replicates.

The melanoma cell line SKmel28 has intermediate IRF4 expression (65 fold HEK293T IRF4 expression) whereas the other melnoma cell line, 501mel has only 2 fold HEK293T IRF4 expression (Figure 6).

4.6 Analysis of the Lymphocyte Specific IRF4 Looping Region Using Epigenetic Data from ENCODE

The ENCODE data of the B-lymphoblastoid cell line GM12878 was utilized to gain a better understanding of the region participating in the chromatin loop to the IRF4 promoter in RPCI-WM1 and U266B1 and to narrow down a region of interest for further testing in RPCI-WM1. GM12878 is an Epstein-Barr Virus transformed B-Lymphocyte cell line, has a relatively normal karyotype and expresses IRF4 mRNA (https://www.genome.gov/26524238). In the cell line GM12878, enhancer predictive data like Dnase hypersensitivity (HS) and ChIP-seq of various transcription factors and histone markers were inspected in the IRF4 looping region (Figure 7). Two adjacent DNase hypersensitivity peaks located at the center of the looping region (16q) overlap with ChIP-seq peaks of several transcription factors (Figure 7) that have known roles in the germinal center reaction in B-cells and plasma cell differentiation.

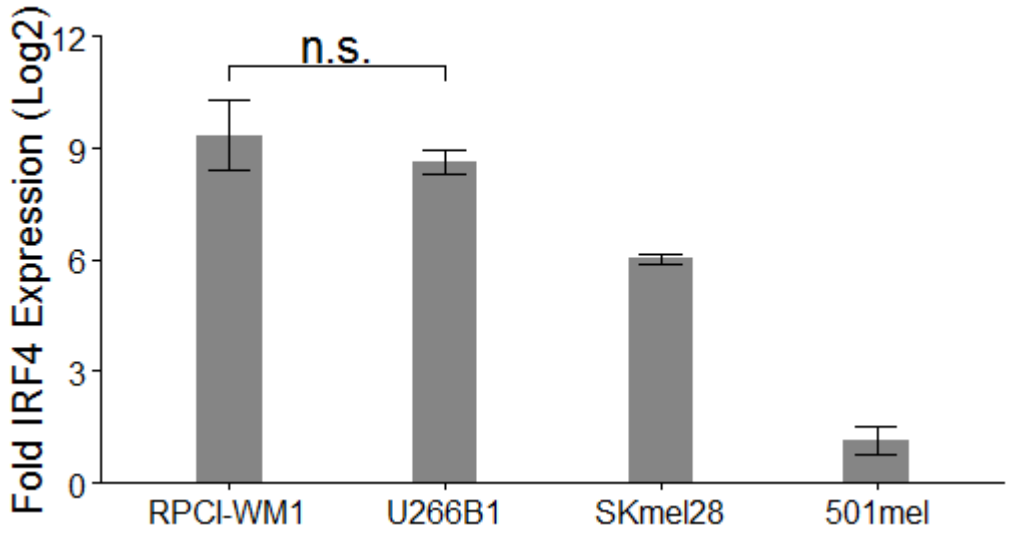

Figure 6. IRF4 mRNA expression

Fold difference of IRF4 expression in cell lines RPCI-WM1, U266B1, SKmel28, 501mel over the IRF4 expression in HEK293T. Error bars correspond to standard deviation between three biological replicates.

This indicates that the region might be under complex regulation during the germinal center transition. The histone markers H3K4me1 and H3K27ac also accumulate around those two DNase hypersensitivity peaks. The 16qE1 region contains an AICE motif which can be bound by heterodimer of IRF4 and the myeloid and B-Cell specific transcription factor PU.1. The two small regions, about 400bp in size and termed 16qE1 and 16qE2 (Figure 7), are marked as potential enhancers. However, based on 1 kb resolution Hi-C data from Rao et al. (2014), no chromatin loops can be seen between the q15, q16, q17 region and the IRF4 promoter in the GM12878 cell line.

4.7 Histone Modification ChIP-qPCR of the IRF4 locus in RPCI-WM1

The potential enhancer activity of the subregions termed 16qE1 and 16qE2 within the 17kb region considered most likely to contain enhancer activity based on data from the ENCODE project underwent further testing. The histone modifications H3K4me1 and H3K27ac have been linked to active regulatory regions as discussed in the Introduction section and have a predictive value on enhancer location in the genome. To examine the histone modifications of the subregions 16qE1 and 16qE2, ChIP-qPCR of the histone modifications H3K4me1 and H3K27ac along with H3K4me3 were carried through in the RPCI-WM1 cell line in two replicates (Method section 3.5). The results are shown in Figure 8.

Both biological replicates of the histone mark ChIP in the RPCI-WM1 cells showed the expected signals (Figure 8). The Myc promoter is predicted to be an active promoter in myeloma cell lines and shows a high H3K4me3 to H3K4me1 ratio, as expected. The reverse is true for the plasma cell enhancers tested (Xbp1E, Blimp1E1 and Blimp1E2) which display predicted signals of high H3K4me1 to H3K4me3 ratio. The negative control Tyrosine intron 4 region shows low signals as expected. It is however unexpected to see the low acetylation signal observed in replicate 2. This might have been because of storage failure of Na-butyrate which is needed to conserve the acetylation signal for successful pulldown of H3K27Ac marked histones.

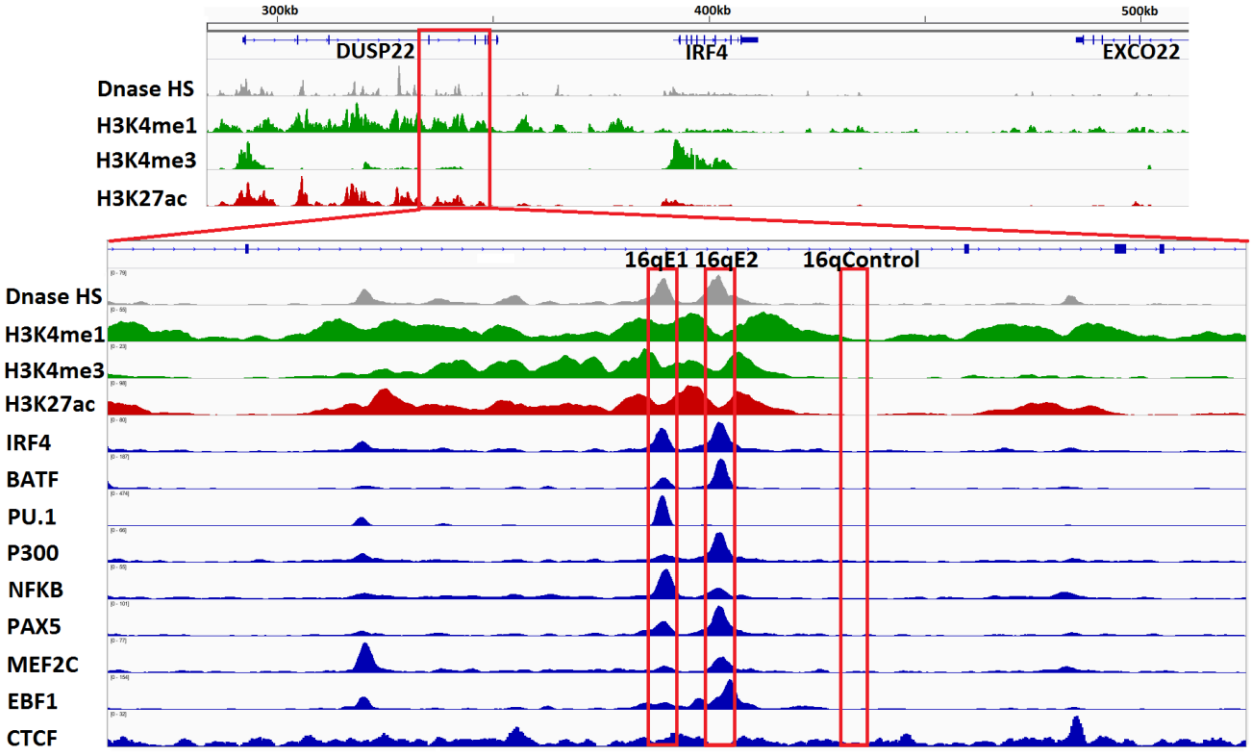

Figure 7. Epigenetic data from ENCODE of the looping region in RPCI-WM1.

DNase seq and ChIP seq for transcription factors IRF4, PU.1, BATF, p300, PAX5, NF-kβ, PAX5, mEF2C, EFBF1 and CTCF and histone markers (H3K4me1, H3K4me3 and H3K27Ac) in the GM12878 cell line. Data from Encode (Consortium, 2012), viewed in IGV (Robinson et al., 2011). Regions tested further and corresponding to potential enhancers (16qE1/2) and a control region (16qControl) are marked with red boxes.

Both ChIP replicates for histone marks were kept and used to check the histone signal profile of the IRF4 locus (Figure 8a and 8b). Primers were designed around active or inactive regions based on Encode histone modification signals in GM12878 to get a rough estimate of the histone modification signals in RPCI-WM1 compared to GM12878. The prediction based on Encode data align modestly with the histone modification ChIP in the RPCI-WM1 cells as demonstrated by fragments 6, 15, 11 and 7 in replicate 1 (Figure 8A). Few odd signals can be seen in replicate 2 (Figure 8B) as the strong monomethylation on fragment 7 and overall low acetylation of the IRF4 promoter and introns which was also seen for the positive controls. Otherwise the two replicates are in good agreement. Because of dilution mistake, 16qE1 and 16qE2 could not be quantified in replicate 1. Thus these regions were only tested in replicate 2. A strong acetylaction signal can be seen for both 16qE1 and 16qE2 but surprisingly for the potential enhancer regions the Histone 3 lysine 4 trimethylation signal is stronger than the monomethylation signal.

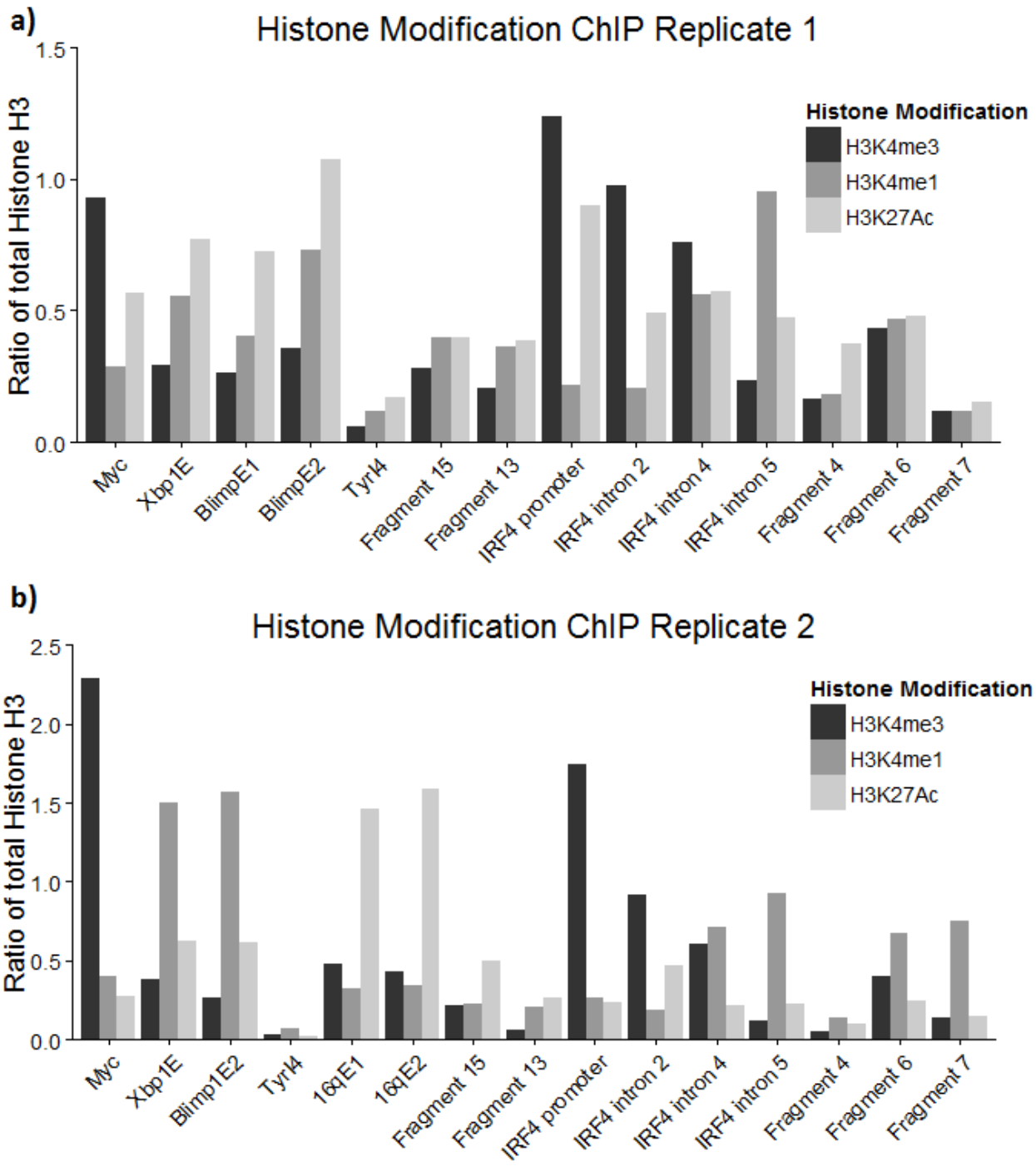

Figure 8. Histone Modification ChIP-qPCR of the IRF4 locus in RPCI-WM1.

Results of histone modification ChIP-qPCR in two replicates, a) and b), of RPCI-WM1 shown as ratio of total Histone H3 for HindIII fragments around IRF4, IRF4 introns and the IRF4 promoter. Control regions; active promoter (Myc), active enhancers (Xbp1 plasma cell enhancer (Xbp1E), Blimp1 plasma cell enhancers (Blimp1E1 and Blimp1E2) and negative control (Tyrosinase intron 4 melanoma enhancer (TyrI4)).

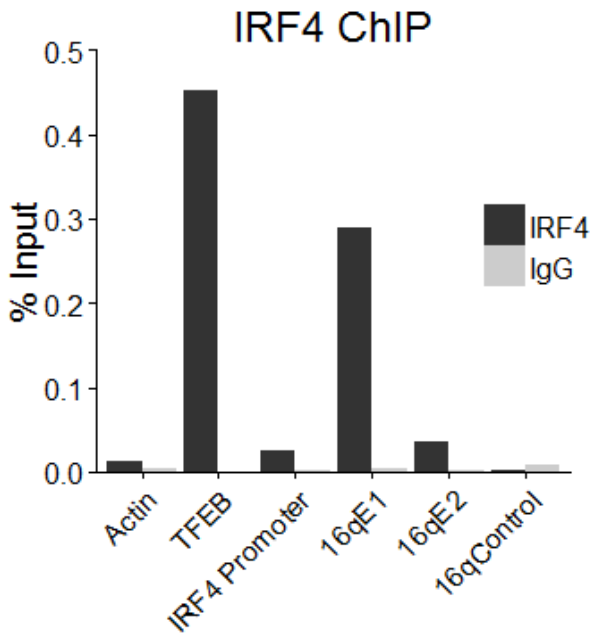

Figure 9. IRF4 ChIP in RPCI-WM1.

Result of IRF4 ChIP-qPCR in RPCI-WM1 cells of actin, TFEB and IRF4 promoters and 16qE1/E2/control as percent of input.

4.8 IRF4 ChIP-qPCR in RPCI-WM1

The IRF4 protein has been predicted to be positively auto-regulated in plasma and myeloma cells (Martinez et al., 2012 and Shaffer et al., 2008) and the binding sites of transcription factors have a predictive value in enhancer detection (Dogan et al., 2015). Therefore the binding of the IRF4 protein to the potential enhancer regions (16qE1/16qE2) was assessed using an IRF4 ChIP in RPCI-WM1 cells (Methods section 3.6) previously made by a Ph.D student in the lab (Dilixiati Remina, unpublished data).

IRF4 binds the promoter of TFEB in GM12878 according to IRF4 ChIP-seq (ENCODE) and works as a positive control for the IRF4 ChIP. A strong IRF4 signal compared to the Immunoglobulin G (IgG) signal at the TFEB promoter and a low IRF4 signal at the actin promoter indicate a successfully optimized IRF4 ChIP (Figure 9). This IRF4 ChIP was then used to test for binding of IRF4 on 16qE1, 16qE2 and 16qControl using qPCR (Figure 9). IRF4 binds strongly to the 16qE1 region and fairly to the IRF4-promoter and the 16qE2 regions but almost no IRF4 is noticeable on the 16qControl region. The IRF4 protein might have a role in the formation of a chromatin loop between the 15q, 16q and 17q fragments to the IRF4 promoter.

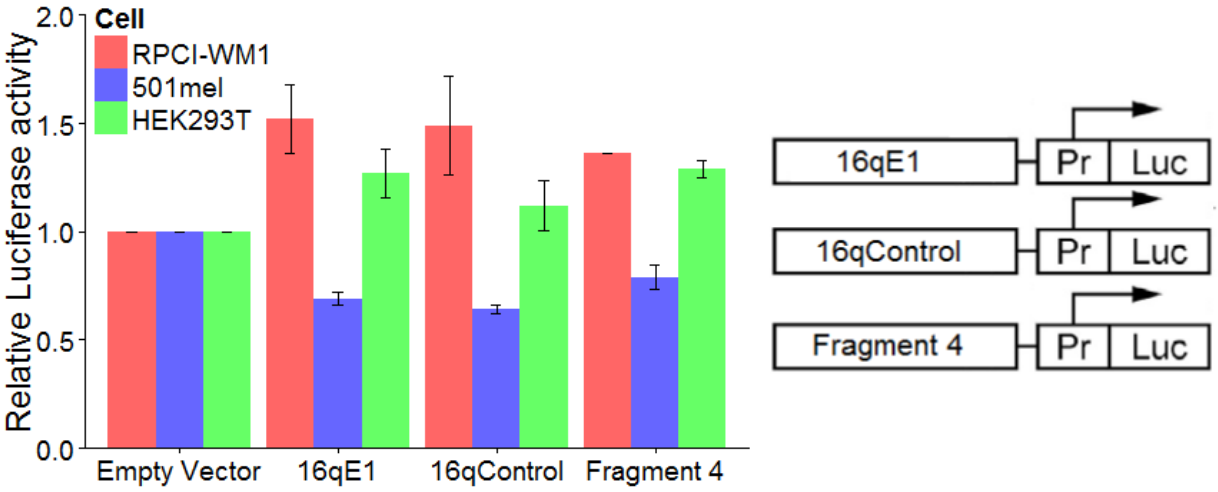

Figure 10. Enhancer activity assessment of the 16qE1 subregion using luciferase assay

Luciferase activity normalized against renilla activity for 16qE1, 16qControl, Fragment 4 and empty pGL3-promoter vector in RPCI-WM1, 501mel and HEK239T cell lines. Error bars correspond to standard deviation. The regions 16qE1, 16qControl, Fragment 4 were cloned in front of the CMV promoter (Pr) and the luciferase gene (Luc) on the pGL3-promoter plasmid.

4.9 Assessment of the Enhancer Activity of the Subregion 16qE1 using transcription activation Assays

Transcription activation assays are a common tool to study the activity of genomic regions. The region of interest is cloned upstream of a luciferase gene and the ability of the region to drive luciferase expression, and therefore its enhancer potential, can be evaluated. The 16qE1 region which contains promising evidence for enhancer activity was tested in such an assay. The 402bp region and the negative control region were cloned into the pGL3-promoter (method section 3.6) to make the vectors pGL3-promoter-16qE1 and pGL3-promoter-16qControl. These two vectors along with the empty vector and the already made pGL3-promoter-fragment4 were used to co-transfect the cell lines RPCI-WM1, 501mel and HEK293T cells with pCMV-renilla in triplicate (method section 3.6). The results are shown in figure 10.

The 16qE1 region seems to cause a modest increase in luciferase activity in the RPCI-WM1 cell line when compared to the empty pGL3-promoter. This pattern of increased luciferase activity over the activity of the empty vector was however also seen in the two negative control pGL3-promoter vectors (16qControl and fragment 4). The same pattern, even similar values, was observed in the HEK293T cell line and in the 501mel cells the luciferase signal was consistently lower for all three vectors compared to the empty pGL3-promoter vector. Thus, the 16qE1 region shows to some extent an increase in enhancer activity when compared to the empty pGL3-promoter but the increase in enhancer activity is similar to the increase caused by control regions which are not expected to contain an active region.

5. Discussion

5.1 Chromosome Conformation Capture set up.

The chromosome conformation capture (3C) technique, a powerful technique to study chromatin interactions, was successfully established in the laboratory. Hematopoietic cell lines such as RPCI-WM1 and U266B1 have been rated excellent in 3C performance compared to the adherent melanocytes and HEK293T cell lines which have been rated poor (Stadhouders et al., 2013). Here the hematopoietic cell lines generally performed better in 3C and only moderate success was seen in the digestion and ligation efficiency of the 3C samples for the adherent cell lines. Further optimization might be needed for the adherent cell lines to increase digestion and ligation efficiency in order to increase 3C performance. Quantification of 3C samples was improved over the course of the study going from 3C-Gel to 3C-qPCR to 3C-ddPCR, progressively increasing the accuracy and reproducibility of the method. One expected pattern of 3C data is that restriction fragments closer on the linear scale should be interacting more frequently than more distant restriction fragments as the number of random collusions decreases with increased distance. This was seen in most cell lines as interaction frequency decreased the further away from the IRF4 promoter the restriction fragment was. With the 3C method and derived techniques we have the potential to find cell specific as well as common long range interactions of the IRF4 promoter and shed light on the molecular mechanism of its regulation and how it might differ in the different cell types.

5.2 IRF4 chromatin loops in RPCI-WM1 and U266B1

It is important to study the transcriptional regulation of IRF4 because of its role in plasma cell development and myeloma. Previously it has been demonstrated that enhancer regions in direct contact with distant promoters can affect the expression of their corresponding genes (Rao et al., 2014). Higher frequency of interactions between the enhancer and the promoter would translate into increased gene expression.

The 3C method was applied to five cell lines: the Waldenström macroglobulinemia cell line RPCI-WM1, the myeloma cell line U266B1, the melanoma cell lines 501mel and SKmel28 and the human embryonic kidney cell line 293T (HEK293T). This revealed a 17kb region upstream of the IRF4 promoter that forms a strong chromatin loop to the IRF4 promoter in the RPCI-WM1 cell line and a weaker loop in the U266B1 and the 501mel cell lines but not the other two cell lines tested. The high chromosome fibre interaction background in 501mel diminishes the potential biological function of the chromatin loop seen between the q14-q16 fragments and the IRF4 promoter in 501mel. On the other hand the chromatin loop seen in the U266B1 cell has higher interaction frequencies than the background chromosome fibre interactions in U266B1. This was demonstrated by a decrease in interaction frequencies seen in fragments q11, q12 and q13 located closer to the IRF4 promoter than the chromatin loop in U266B1. This might be indicative of a B cell linage specific interaction between the IRF4 promoter and this 17kb region corresponding to the *HindIII* fragments q15, q16 and q17. In agreement, IRF4 expression is significantly higher in the U266B1 and RPCI-WM1 cell lines than in the melanoma cell lines. However, this B cell linage specific interaction does not tell the whole story of IRF4 regulation as

there is still a noticeable IRF4 expression in SKmel28 and a significant difference in IRF4 expression between RPCI-WM1 and U266B1 was not seen even though the interaction is significantly stronger in RPCI-WM1. Indicating another regulatory looping events outside of the regions tested here or a difference in the transcription factor network in U266B1 and RPCI-WM1 at the IRF4 promoter.

The ENCODE data for the B-lymphoblastoid cell line GM12878 was utilized to examine the 17kb looping region and identify potential enhancer regions within the looping region, looking mainly at IRF4 and p300 binding, transcription factors that have role in the germinal centers, as well as DNase I hypersensitivity pattern and histone modification peaks (H3K4me1, H3K4me3 and H3K27ac). Two subregions (16qE1/E2) within the 17kb looping region were considered most likely to hold enhancer activity based on the epigenetic data from the ENCODE project. Accumulation of histone modifications associated with enhancer regions was seen around both regions, as expected for the potential enhancer regions. This signature of histone modifications was confirmed for both regions with a histone modification according to ChIP-qPCR experiments performed on the RPCI-WM1 cell line. However the H3K4me1 to H3K4me3 signal ratio was weaker than expected as it is believed that enhancer regions display a strong H3K4me1 to H3K4me3 signal ratio.

The Encode data from the B cell line GM12878 suggests two different IRF4 binding sites at the looping region. One is a binding of PU.1-IRF4 heterodimer to an EICE motif (16qE1) whereas the other is an IRF4 heterodimer with BATF binding to an AICE motif (16qE2). IRF4 can bind these motifs in low concentration unlike the IRES motifs which IRF4 homodimer bind with low affinity and thus requires higher cellular concentration of IRF4. IRF4 can bind to regions containing EICE or AICE motif in native B lymphocytes with low IRF4 concentration (Ochiai et al 2013). Accordingly the putative enhancer region might be a weak enhancer upregulating IRF4 in B lymphocytes in response to a B cell receptor activation and an entry into the germinal centre reactions. The binding of the IRF4 protein was confirmed for only one of the subregions (16qE1). This binding of the IRF4 protein to the looping region raises the question if IRF4 expression is under a positive auto-regulation. A positive auto-regulation of IRF4 in the context of plasma cell differentiation has been suggested (Martinez et al., 2012; Shaffer et al., 2008). Low levels of IRF4 expression keep B lymphocytes within the germinal center reactions but increase in IRF4 expression pushes the cell towards the plasma cell fate. This positive auto-regulation of IRF4 provides a mechanism to prevent dedifferentiation (Martinez et al., 2012).

Finally, the ability of the more promising subsequence, 16qE1, to drive luciferase expression in a transcription activation assay experiment was tested, revealing only modest enhancer activity of the 16qE1 subsequence compared to empty vector and no activity compared to control regions. However, negative or modest results in luciferase assays can be hard to interpret and there may be many reasons why it failed. For example, inclusion of inhibiting motif in the region tested, E. coli directed DNA methylation on the region or even unexpected enhancer activity of negative controls might be reasons for negative luciferase assay results.

Absence of chromatin looping between the B cell linage specific enhancer region and the IRF4 promoter in the GM12878 cell line is indicative of an important difference of the region in the different cell lines, however the ENCODE data for the GM12878 cell line predicts an enhancer at the same region. Indeed one can assume that even though the B lymphoblastoid cell line GM12878 expresses IRF4 the

expression is most likely lower than in both U266B1 and RPCI-WM1. IRF4 has been shown to be vital for the survival of myeloma cell lines as IRF4 controls a fused B lymphocyte and plasma cell transcription factor network. The q15, q16 and q17 looping region could therefore be a B cell linage specific enhancer region aberrantly taking part in the upregulation of IRF4 in the RPCI-WM1 and U266B1 cell lines.

5.3 IRF4 chromatin loops in 501mel and SKmel28

The two melanoma cell lines tested, SKmel28 and 501mel showed elevated chromatin fibre interaction between the nearest *HindIII* fragment tested, q10 and the IRF4 promoter. Because of the linear distance and the resolution of a 3C with a 6-base cutter, no isolated peak of interaction can be seen. Therefore the elevated interaction as a cause of random interactions between the q10 fragment and the IRF4 promoter cannot be ruled out. On the other hand the fact that the increased interaction was only seen in the melanoma cell lines raises a question if this interaction has a biological function. In fact the q10 region might be having a repressing effect on the IRF4 expression as the interaction between the q10 region and the IRF4 promoter is very strong in the 501mel cell line which has very low IRF4 expression compared to the other melanoma cell line SKmel28.

5.4 Future directions

The location of the looping region in RPCI-WM1 identified with the 3C experiment needs to be tested further. It would be logical to rule out any bias that might occur cause of apparent regional digestion by *HindIII* by using another restriction enzyme and thus confirm the chromatin looping of the region (15q, 16q and 17q) to the IRF4 promoter. Additional screening of the looping region would involve a 4-base restriction enzyme that would allow for higher resolution of the looping. The restriction fragments of a 4-base restriction enzyme are generally hundreds of base pairs long compared to the fragments made by 6-base restriction enzyme such as *HindIII* which are thousands of base pairs long.

In this study we only investigated a region 96kb upstream and a 126kb region downstream of the IRF4 promoter using 3C to find long range interactions and in turn potential enhancer regions of the IRF4 promoter. This study does not rule out more chromatin looping events on the IRF4 promoter that might be taking place outside of the tested region as promoters can participate in multiple looping events. This could be done with the 4C method, having the benefit of screening all looping regions. 4C either done from the IRF4-promoter (finding other potential enhancers that regulate IRF4 expression) or from the 17kb looping region in RPCI-WM1 and U266B1 (possibly finding other genes under this enhancer's control) may be useful for identifying novel interactions.

A more time and cost intensive but at the same time an improved method to show enhancer activity of a region would be to knock it out or mutate motifs of key transcription factors within the region of interest using a genome editing tool like CRISPR. A decrease in target gene expression, IRF4, would then hint towards an active enhancer in the region. Additionally, the transcription of enhancers generating eRNA has been suggested to have a functional role in enhancer activity. A decrease in IRF4 expression following siRNA mediated knock down of eRNAs in the 16qE1 region could therefore reveal an enhancer activity of the region.

The context depended role of IRF4 in plasma cell differentiation is intriguing and the concentration of IRF4 has been shown to affect the binding of IRF4 to different motifs (Ochiai et al., 2013). In this context it would be interesting to investigate the changes in chromatin looping of the IRF4 locus during plasma cell development and possibly in a cell system with controlled IRF4 expression. Observing the changes in chromatin looping going from low to high IFR4 concentration as the difference in IRF4 expression is known to direct cell fate.

References

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., . . . Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature, 507*(7493), 455-461.

Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M., & Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science, 339*(6123), 1074-1077.

Bairoch, A. (2016). The Cellosaurus: a cell line knowledge resource. Retrieved from http://web.expasy.org/cellosaurus/

Bickmore, W. A. (2013). The spatial organization of the human genome. *Annu Rev Genomics Hum Genet, 14*, 67-84.

Birnbaum, R. Y., Patwardhan, R. P., Kim, M. J., Findlay, G. M., Martin, B., Zhao, J., . . . Ahituv, N. (2014). Systematic dissection of coding exons at single nucleotide resolution supports an additional role in cell-specific transcriptional regulation. *PLoS Genet, 10*(10), e1004592.

Branco, M. R., & Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol, 4*(5), e138.

Calo, E., & Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Mol Cell, 49*(5), 825-837.

Carrasco, D. R., Sukhdeo, K., Protopopova, M., Sinha, R., Enos, M., Carrasco, D. E., . . . DePinho, R. A. (2007). The differentiation and stress response factor XBP-1 drives multiple myeloma pathogenesis. *Cancer Cell, 11*(4), 349-360.

Chitta, K. S., Paulus, A., Ailawadhi, S., Foster, B. A., Moser, M. T., Starostik, P., . . . Chanan-Khan, A. A. (2013). Development and characterization of a novel human Waldenstrom macroglobulinemia cell line: RPCI-WM1, Roswell Park Cancer Institute - Waldenstrom Macroglobulinemia 1. *Leuk Lymphoma, 54*(2), 387-396.

Chung, J. B., Silverman, M., & Monroe, J. G. (2003). Transitional B cells: step by step towards immune competence. *Trends Immunol, 24*(6), 343-349.

Cichorek, M., Wachulska, M., Stasiewicz, A., & Tyminska, A. (2013). Skin melanocytes: biology and development. *Postepy Dermatol Alergol, 30*(1), 30-41.

Consortium, E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), 57-74.

De Silva, N. S., & Klein, U. (2015). Dynamics of B cells in germinal centres. *Nat Rev Immunol, 15*(3), 137-148.

De Silva, N. S., Simonetti, G., Heise, N., & Klein, U. (2012). The diverse roles of IRF4 in late germinal center B-cell differentiation. *Immunol Rev, 247*(1), 73-92.

de Wit, E., & de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev, 26*(1), 11-24.

Dekker, J., & Misteli, T. (2015). Long-Range Chromatin Interactions. *Cold Spring Harb Perspect Biol, 7*(10), a019356.

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science, 295*(5558), 1306-1311.

Deng, B., Melnik, S., & Cook, P. R. (2013). Transcription factories, chromatin loops, and the dysregulation of gene expression in malignancy. *Semin Cancer Biol, 23*(2), 65-71.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., . . . Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature, 485*(7398), 376-380.

Do, T. N., Ucisik-Akkaya, E., Davis, C. F., Morrison, B. A., & Dorak, M. T. (2010). An intronic polymorphism of IRF4 gene influences gene transcription in vitro and shows a risk association with childhood acute lymphoblastic leukemia in males. *Biochim Biophys Acta, 1802*(2), 292-300.

Dogan, N., Wu, W., Morrissey, C. S., Chen, K. B., Stonestrom, A., Long, M., . . . Hardison, R. C. (2015). Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin, 8*, 16.

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., . . . Dekker, J. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res, 16*(10), 1299-1309.

Gavrilov, A., Eivazova, E., Priozhkova, I., Lipinski, M., Razin, S., & Vassetzky, Y. (2009). Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. *Methods Mol Biol, 567*, 171-188.

Gibcus, J. H., & Dekker, J. (2013). The hierarchy of the 3D genome. *Mol Cell, 49*(5), 773-782.

Grossman, A., Mittrucker, H. W., Nicholl, J., Suzuki, A., Chung, S., Antonio, L., . . . Mak, T. W. (1996). Cloning of human lymphocyte-specific interferon regulatory factor (hLSIRF/hIRF4) and mapping of the gene to 6p23-p25. *Genomics, 37*(2), 229-233.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., . . . Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell, 162*(4), 900-910.

Hagege, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., . . . Forne, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc, 2*(7), 1722-1733.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell, 144*(5), 646-674.

Hardison, R. C., & Taylor, J. (2012). Genomic approaches towards finding cis-regulatory modules in animals. *Nat Rev Genet, 13*(7), 469-483.

Heidari, N., Phanstiel, D. H., He, C., Grubert, F., Jahanbani, F., Kasowski, M., . . . Snyder, M. P. (2014). Genome-wide map of regulatory interactions in the human genome. *Genome Res, 24*(12), 1905-1917.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., . . . Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature, 459*(7243), 108-112.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A, 106*(23), 9362-9367.

Hu, C. C., Dougan, S. K., McGehee, A. M., Love, J. C., & Ploegh, H. L. (2009). XBP-1 regulates signal transduction, transcription factors and bone marrow colonization in B cells. *EMBO J, 28*(11), 1624-1636.

Inoue, F., & Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. *Genomics, 106*(3), 159-164.

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., . . . Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature, 503*(7475), 290-294.

Jin, Q., Yu, L. R., Wang, L., Zhang, Z., Kasper, L. H., Lee, J. E., . . . Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J, 30*(2), 249-262.

Jónasson, J. G., & Tryggvadóttir, L. (2012). *Krabbamein á Íslandi - Upplýsingar úr Krabbameinsskrá fyrir tímabilið 1955-2010* (Krabbameinsfélagið Ed.). Reykjavík Krabbameinsfélagið.

Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., . . . Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res, 23*(5), 800-811.

Kitano, M., Moriyama, S., Ando, Y., Hikida, M., Mori, Y., Kurosaki, T., & Okada, T. (2011). Bcl6 protein expression shapes pre-germinal center B cell dynamics and follicular helper T cell heterogeneity. *Immunity, 34*(6), 961-972.

Klein, U., Casola, S., Cattoretti, G., Shen, Q., Lia, M., Mo, T., . . . Dalla-Favera, R. (2006). Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination. *Nat Immunol, 7*(7), 773-782.

Kwasnieski, J. C., Fiore, C., Chaudhari, H. G., & Cohen, B. A. (2014). High-throughput functional testing of ENCODE segmentation predictions. *Genome Res, 24*(10), 1595-1602.

Kwasnieski, J. C., Mogno, I., Myers, C. A., Corbo, J. C., & Cohen, B. A. (2012). Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc Natl Acad Sci U S A, 109*(47), 19498-19503.

Lam, M. T., Cho, H., Lesch, H. P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., . . . Glass, C. K. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature, 498*(7455), 511-515.

Lee, T. I., Johnstone, S. E., & Young, R. A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc, 1*(2), 729-748.

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., . . . Ruan, Y. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell, 148*(1-2), 84-98.

Li, S., & Ovcharenko, I. (2015). Human Enhancers Are Fragile and Prone to Deactivating Mutations. *Mol Biol Evol, 32*(8), 2161-2180.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., . . . Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science, 326*(5950), 289-293.

Lin, L., Gerth, A. J., & Peng, S. L. (2004). Active inhibition of plasma cell development in resting B cells by microphthalmia-associated transcription factor. *J Exp Med, 200*(1), 115-122.
Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., & Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell, 126*(2), 403-413.
Markova, E. N., Kantidze, O. L., & Razin, S. V. (2011). Transcriptional regulation and spatial organisation of the human AML1/RUNX1 gene. *J Cell Biochem, 112*(8), 1997-2005.
Martinez, M. R., Corradin, A., Klein, U., Alvarez, M. J., Toffolo, G. M., di Camillo, B., . . . Stolovitzky, G. A. (2012). Quantitative modeling of the terminal differentiation of B cells and mechanisms of lymphomagenesis. *Proc Natl Acad Sci U S A, 109*(7), 2672-2677.
Melo, C. A., Drost, J., Wijchers, P. J., van de Werken, H., de Wit, E., Oude Vrielink, J. A., . . . Agami, R. (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell, 49*(3), 524-535.
Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., . . . Osborne, C. S. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet, 47*(6), 598-606.
Mittrucker, H. W., Matsuyama, T., Grossman, A., Kundig, T. M., Potter, J., Shahinian, A., . . . Mak, T. W. (1997). Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. *Science, 275*(5299), 540-543.
Morgan, G. J., Walker, B. A., & Davies, F. E. (2012). The genetic architecture of multiple myeloma. *Nat Rev Cancer, 12*(5), 335-348.
Natkunam, Y., Warnke, R. A., Montgomery, K., Falini, B., & van De Rijn, M. (2001). Analysis of MUM1/IRF4 protein expression using tissue microarrays and immunohistochemistry. *Mod Pathol, 14*(7), 686-694.
Natoli, G., & Andrau, J. C. (2012). Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet, 46*, 1-19.
Naumova, N., Smith, E. M., Zhan, Y., & Dekker, J. (2012). Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods, 58*(3), 192-203.
Nutt, S. L., Hodgkin, P. D., Tarlinton, D. M., & Corcoran, L. M. (2015). The generation of antibody-secreting plasma cells. *Nat Rev Immunol, 15*(3), 160-171.
Ochiai, K., Maienschein-Cline, M., Simonetti, G., Chen, J., Rosenthal, R., Brink, R., . . . Sciammas, R. (2013). Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4. *Immunity, 38*(5), 918-929.
Parekh, S., Polo, J. M., Shaknovich, R., Juszczynski, P., Lev, P., Ranuncolo, S. M., . . . Melnick, A. (2007). BCL6 programs lymphoma cells for survival and differentiation through distinct biochemical mechanisms. *Blood, 110*(6), 2067-2074.
Pasqualucci, L., Compagno, M., Houldsworth, J., Monti, S., Grunn, A., Nandula, S. V., . . . Dalla-Favera, R. (2006). Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J Exp Med, 203*(2), 311-317.
Phillips, J. E., & Corces, V. G. (2009). CTCF: master weaver of the genome. *Cell, 137*(7), 1194-1211.
Praetorius, C., Grill, C., Stacey, S. N., Metcalf, A. M., Gorkin, D. U., Robinson, K. C., . . . Steingrimsson, E. (2013). A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell, 155*(5), 1022-1033.
Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature, 470*(7333), 279-283.
Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., . . . Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell, 159*(7), 1665-1680.
Recaldin, T., & Fear, D. J. (2016). Transcription factors regulating B cell fate in the germinal centre. *Clin Exp Immunol, 183*(1), 65-75.
Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nat Biotechnol, 29*(1), 24-26.
Roe., B. (1999). DNA Isolation by a Cleared Lysate Method Followed by Double Acetate Precipitation Version 3b. Retrieved from http://www.genome.ou.edu/DblAcetateProcV3.html
Rowley, M. J., & Corces, V. G. (2016). The three-dimensional genome: principles and roles of long-distance interactions. *Curr Opin Cell Biol, 40*, 8-14.
Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature, 489*(7414), 109-113.

Schebesta, A., McManus, S., Salvagiotto, G., Delogu, A., Busslinger, G. A., & Busslinger, M. (2007). Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration, and immune function. *Immunity, 27*(1), 49-63.

Sciammas, R., Shaffer, A. L., Schatz, J. H., Zhao, H., Staudt, L. M., & Singh, H. (2006). Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation. *Immunity, 25*(2), 225-236.

Shaffer, A. L., Emre, N. C., Lamy, L., Ngo, V. N., Wright, G., Xiao, W., . . . Staudt, L. M. (2008). IRF4 addiction in multiple myeloma. *Nature, 454*(7201), 226-231.

Shaffer, A. L., Lin, K. I., Kuo, T. C., Yu, X., Hurt, E. M., Rosenwald, A., . . . Staudt, L. M. (2002). Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity, 17*(1), 51-62.

Shapiro-Shelef, M., & Calame, K. (2005). Regulation of plasma-cell development. *Nat Rev Immunol, 5*(3), 230-242.

Shapiro-Shelef, M., Lin, K. I., McHeyzer-Williams, L. J., Liao, J., McHeyzer-Williams, M. G., & Calame, K. (2003). Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity, 19*(4), 607-620.

Solimini, N. L., Luo, J., & Elledge, S. J. (2007). Non-oncogene addiction and the stress phenotype of cancer cells. *Cell, 130*(6), 986-988.

Stadhouders, R., Kolovos, P., Brouwer, R., Zuin, J., van den Heuvel, A., Kockx, C., . . . Soler, E. (2013). Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc, 8*(3), 509-524.

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., . . . Stefansson, K. (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet, 39*(12), 1443-1452.

Tang, Z., Luo, O. J., Li, X., Zheng, M., Zhu, J. J., Szalaj, P., . . . Ruan, Y. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell, 163*(7), 1611-1627.

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., . . . Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature, 489*(7414), 75-82.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell, 10*(6), 1453-1465.

Travers, P., Walport, M., Schlomchik, M., & Janeway, C. (2001). *Immunobiology: The Immune System in Health and Disease* (Vol. 5th). New York: Garland Publishing.

van Arensbergen, J., van Steensel, B., & Bussemaker, H. J. (2014). In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol, 24*(11), 695-702.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., . . . Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature, 457*(7231), 854-858.

Visser, M., Palstra, R. J., & Kayser, M. (2015). Allele-specific transcriptional regulation of IRF4 in melanocytes is mediated by chromatin looping of the intronic rs12203592 enhancer to the IRF4 promoter. *Hum Mol Genet, 24*(9), 2649-2661.

Walsh, S. H., Laurell, A., Sundstrom, G., Roos, G., Sundstrom, C., & Rosenquist, R. (2005). Lymphoplasmacytic lymphoma/Waldenstrom's macroglobulinemia derives from an extensively hypermutated B cell that lacks ongoing somatic hypermutation. *Leuk Res, 29*(7), 729-734.

Willis, S. N., Good-Jacobson, K. L., Curtis, J., Light, A., Tellier, J., Shi, W., . . . Nutt, S. L. (2014). Transcription factor IRF4 regulates germinal center cell formation through a B cell-intrinsic mechanism. *J Immunol, 192*(7), 3200-3206.

Xu, H., Chaudhri, V. K., Wu, Z., Biliouris, K., Dienger-Stambaugh, K., Rochman, Y., & Singh, H. (2015). Regulation of bifurcating B cell trajectories by mutual antagonism between transcription factors IRF4 and IRF8. *Nat Immunol, 16*(12), 1274-1281.

Zhang, Y., McCord, R. P., Ho, Y. J., Lajoie, B. R., Hildebrand, D. G., Simon, A. C., . . . Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell, 148*(5), 908-921.

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., . . . Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet, 38*(11), 1341-1347.

Zheng, J. F., Dong, S. S., Wang, Q., Pan, J. L., Chen, S. N., & Qiu, H. Y. (2013). [Deletions and rearrangements of PAX5 gene in B-lineage acute lymphoblastic leukemia]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi, 30*(5), 549-552.