Chapter 1

Introduction

The nucleus is a highly functional organelle, responsible for a huge variety of complex cellular processes. To achieve this, the contents of the nucleus are organised into sub-nuclear compartments which specialise in processes such as transcription and replication. In this chapter I will discuss the current understanding of nuclear structure and organisation, how we have obtained this knowledge, and how it may play a part in the initiation of certain cancers.

1.1 Chromatin

Chromatin is a general term used to describe the DNA packaged in the plethora of proteins that bind to it. There are two principle forms of chromatin: *euchromatin* and *heterochromatin*. These were first described in the early twentieth century due to their differential staining within the nucleus [1].

Heterochromatin stains darkly because it remains highly condensed during interphase, typically relocating to the nuclear periphery. It can be split into *constitutive heterochromatin* and *facultative heterochromatin*; the former describing heterochromatin found in all cell types, typically containing repetitive elements found in centromeres and telomeres; the latter describing less compacted inactive chromatin which can vary between cell types as they differentiate. Heterochromatin is well known as being a repressive environment for gene expression; studies whereby normally active regions are artificially tethered to the inner nuclear membrane show the ablation of gene expression [2, 3] though this effect appears to be locus specific and is not always observed [4]. Euchromatin is the site of most geneic transcription, as shown by the presence of RNA Polymerase and nascent transcriptions found in early fractionation studies [5]. Its looser compaction allows access to the DNA by the cellular machinery; enabling the binding of transcription factors and the initiation of transcription.

The differences between heterochromatin and euchromatin lie within the proteins that they contain. Chromatin

acts as a platform for proteins to bind to, differences in histone modifications, variants, nucleosome packing and DNA modifications affect the accessibility and binding profile of the chromatin, and so how the DNA is interpreted by the cell.

1.1.1 Histones

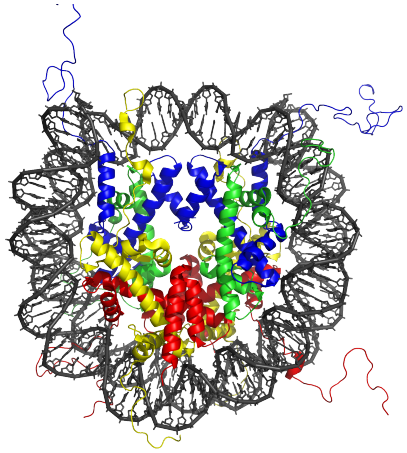

Figure 1.1 – Structure of the nucleosome. X-ray structure of a nucleosome core particle at a resolution of 1.9Å. DNA can be seen wrapped around the core histones, which have flexible tails extending into the nuclear matrix. PDB structure 1KX5 [6].

To package DNA, the double helix is wrapped around an octomer of core histones: two H2A, two H2B, two H3 and two H4. 146 base pairs of DNA wrap around these positively charged proteins to form the nucleosome, which is then bound by histone H1 with linker DNA to make a total of 166 base pairs [6]. This packing forms the 10 nm fibre, often known as 'beads on a string' due to its appearance in electron micrographs. At their most basic level, histones function to compact DNA by counteracting the negative charge of the phosphorylated back bone.

1.1.2 Histone modifications

Core histones have flexible tails which extend outside of the nucleosome, and are accessible to proteins within the nuclear matrix. These tails can be post-translationally modified at a large number of residues and these modifications can affect the packing of chromatin [7] as well as which proteins can bind. The large number of combinatorial possibilities that result result from these modifications have been dubbed the 'Histone Code' [8], different modifications are related to different chromatin states. For example, active promoters are typically enriched for di- and tri-methylation of histone 3 lysine 4 (H3K4) whilst inactive promoters are enriched for H3K27me3 and H3K9me3 (Fig 1.2, for review, see [9]).

As chromatin immunoprecipitation (ChIP) has become a common laboratory technique, combined with microarray techniques (ChIP on chip) and next generation sequencing (ChIP-Seq), our understanding of how histone modifications affect chromatin biology on a genome-wide scale has advanced dramatically. Profiling chromatin types using multiple datasets covering a large number of histone modifications is sufficient to predict the identity and function of regions in the genome with a high degree of accuracy, revealing previously unknown enhancers [10, 11, 12]. Ernst *et al.* used the genome-wide profiles of 9 histone modifications in 9 different cell types to define 15 different chromatin states, describing different states of promoters, enhancers and insulators, amongst others[13]. They integrated data

from genome-wide association studies (GWAS) and found numerous enhancer elements which coincide with disease associated mutations. This systems biology approach can reveal the dynamics across different cell types and can be a powerful tool in understanding how the genome is interpreted in health and disease (see Section 1.7).

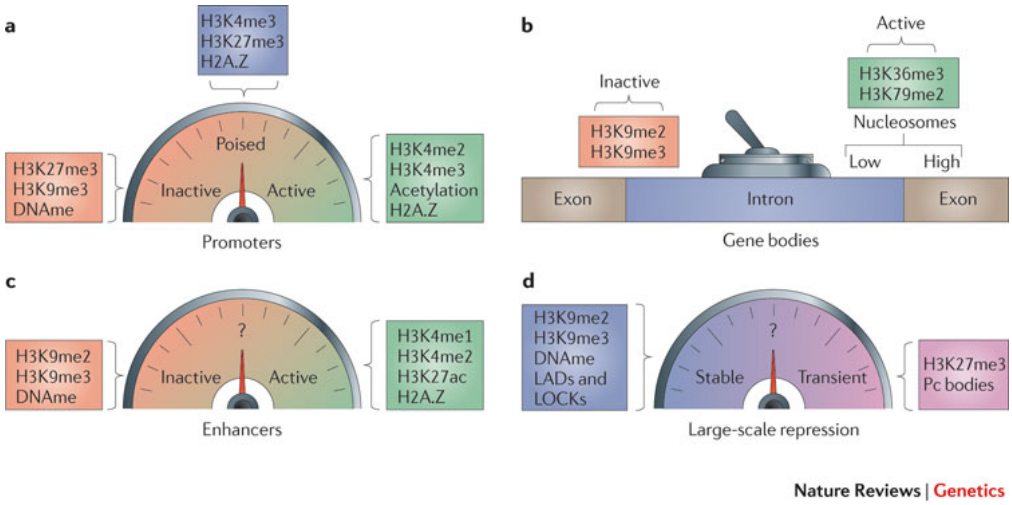

Figure 1.2 – Histone modification associations in chromatin. Diagram showing the range of histone modifications available within mammalian chromatin and how they can specify the behaviour of elements within the DNA. Taken from[9].

1.1.3 Histone variants

In addition to histone tail modifications, chromatin can be modified by the incorporation of histone variants. Canonical core histone genes are found in clustered repeat arrays within the genome, are transcribed during replication and are highly conserved between species. Histone variants are found as single genes spread through the genome, and are subject to far greater diversity [14].

CENP-A is a human variant of histone H3 which replaces the canonical histone in centromeric heterochromatin. It is a key factor in the establishment of the centromeres and kinetochores required for mitosis. The histone variant is incorporated with the help of a number of chaperone proteins, including HJURP, after replication of DNA has finished [15, 16]. CENP-A is essential for the formation of centromeres.

Another frequent histone variant found in humans is H3.3, which differs from canonical H3 by just four amino acids [14]. This histone variant is found within transcribed genes, promoters and regulatory elements, and is thought to be laid down during transcriptional elongation [17]. Nucleosomes containing H3.3 appear to be less stable than canonical nucleosomes, with a high turnover [17]. It is possible that this increased turnover of the nucleosomal components helps to keep the chromatin open and accessible to the transcriptional machinery [14].

Other core histone proteins also have variants, such as H2A.Z, a histone variant found on either side of the nucleosome free regions found at the transcriptional start sites of active genes as well as insulator regions [18]. H2A.Z

promotes the recruitment of RNA Polymerase II by mediating protein interactions [19].

1.2 Two-dimensional organisation

Since the inception of cytogenetics, it has been known that not all chromosomes are the same. The development of chromosome banding techniques in the 1970s allowed detailed human karyotypes to be determined, complete with differential staining of regions within each chromosome [20]. Banding assays can show regions of heterochromatin and euchromatin, highlighting the variation in characteristics across regions of the genome [21].

Put in stuff about yeast and drosophila gene clustering here [22, 23]

The sequencing of the human genome allowed detailed analysis of GC content, gene density and repetitive sequence content [24]. Versteeg *et al.* integrated a multitude of SAGE tag expression profiles from different cell types into the genomic map and built on earlier work defined regions of high transcriptional activity, called ridges [25, 26]. They found ridges to be gene-dense, highly transcribed, have a high GC content and low LINE repeat density. These features are based purely on the underlying sequence content and so do not vary amongst cell types. The different ridges and anti-ridges were found to contain different classes of genes, with weakly expressed genes clustering within anti-ridges and clusters of highly expressed housekeeping genes found predominantly in ridges [25].

Whilst the two-dimensional organisation of the genome cannot completely explain the degree of complexity found within the transcriptome, these studies demonstrate that the order of sequence within the genome is not entirely random, and can affect the transcriptional control of genes.

1.3 Chromosome Territories

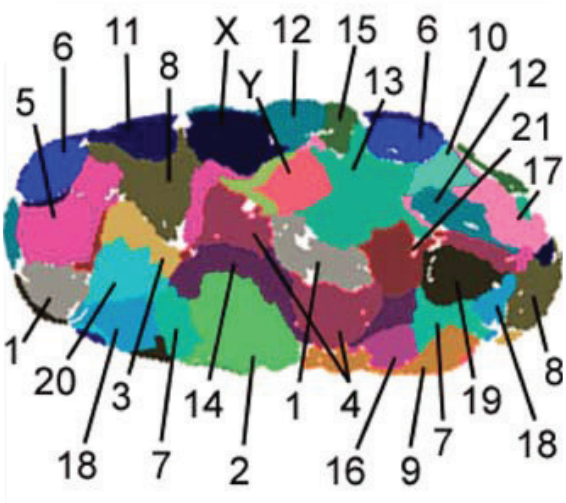

Figure 1.3 – **Chromosome Territories.** Image of a Human fibroblast cell in G0 with all chromosomes labelled using multi-colour FISH. Adapted from Bolzer *et al.* [33]

The interphase nucleus is a highly structured organelle. As chromosomes decondense after metaphase they retain some degree of structure, forming "chromosome territories" (CTs) [27]. Circumstantial evidence for interphase organisation of chromosomes has existed for a long time, first suggested by Carl Rabl in 1885 [28]. Oberservations by Stack *et al.* using microscopy with giemsa-band staining suggested that chromosomes retained some degree of organisation during interphase [29], and in 1982 Cremer *et al.,* showed that interphase chromosomes exist in territories by studying the pattern

of DNA damage in metaphase chromosomes after spot irradiation in during interphase [30]. The subsequent development of chromosome paints, a method to visualise entire- or part- chromosomes with fluorescence *in-situ* hybridisation (FISH), confirmed these findings [31, 32, 33].

As FISH techniques have developed, so too has the detail with which CT organisation can be studied. Several groups have shown that chromosome territory position within the nucleus is not random and correlates with chromosome size [34, 35, 33], gene-density [35, 36] and replication timing [37, 38] - those near the centre of the nucleus tend to be gene-rich, early replicating and small. The specificity of CT positioning is a topic of debate, though studies have suggested conservation of CT positioning through evolution [39] and shown tissue specificity in chromosome pairing [40, 41].

1.3.1 CTs and transcription

It's known that for some but not all genes, positioning at the nuclear periphery correlates with reduced gene expression [42, 43, 44]. To investigate nuclear positioning can cause changes in transcriptional activity or is simply a consequence, two groups published studies which artificially tethered genomic regions to the inner nuclear membrane using Lac operators *(lacO)* [2, 3]. Both groups observed a decrease in the transcriptional activity of the regions when tethered, an effect that was ablated when cells were treated with trichostatin A (TSA) to inhibit class I and class II histone deacetylases [2]. It should be noted that this effect does not appear to apply to all loci [4]. Whilst these studies suggest that it is nuclear positioning that leads to transcriptional effects, a study by Croft *et al.* showed that inhibition of transcription causes a reversible change in CT position [36]. Large scale rearrangements of CTs have also been observed during cell differentiation [45, 46], likely due to changes in transcriptional profiles and chromatin remodelling.

This evidence suggests that there is a dynamic interplay between CT positioning and transcription - gross transcriptional patterns may drive the position of chromosomes within the nucleus, and those positions may in turn affect the transcription of the genes they contain.

1.3.2 Chromosome territory dynamics

The movement of chromatin loci within the nucleus is known to be largely due to local brownian motion, or "constrained diffusion" and seems to be limited by attachment to a nuclear cytoskeleton, nucleoli and the inner nuclear membrane [47, 48]. Other studies have shown a mixture of local diffusion and larger, active movements [49]. Chuang *et al.* studied Chinese hamster ovary (CHO) cells after stimulation with a transcriptional activator and analysed the movement of the target chromosome from the nuclear periphery to the interior of the nucleus [50]. They found that chromosome

movements happened in rapid unidirectional bursts, suggesting an active mechanism. This was supported by a later study in human fibroblast cells by Mehta *et al.* who demonstrated chromosome movement only 15 minutes after serum starvation, a process that was ablated by the inhibition of actin and myosin polymerisation [51].

1.3.3 Chromatin decondensation and the inter-chromosomal space

In addition to the study of whole chromosome positioning, there has been a great deal of research into the positioning of individual sequences relative to their chromosome territory. Early studies suggested that transcribed genes were found at the surface of chromosome territories [52] leading to a model whereby transcriptionally inactive genes are buried within territories and expressed genes contact in an inter-chromosome domain (ICD) where they can contact the transcriptional machinery [53]. This model gained support due to studies showing genes at the periphery of CTs [54] and new techniques to visualise the ICD using microscopy [55]. A number of FISH studies showed genes moving away from their territories in large loops upon activation [56, 57, 58], suggesting that they may be recruited to the ICD for transcription.

Despite these case studies, looping is not a prerequisite for transcription; DNA-FISH studies have shown transcription within the volume of chromosome territories [59, 60]. Osborne *et al.* showed that actively transcribed *Uros* gene is more frequently outside it's CT than the inactive gene *Fgfr2*, though this position was not necessary for transcription, suggesting that actively transcribed genes may preferentially locate to the surface of CTs, but this alone is not sufficient to drive transcription [61]. As such, it has been suggested that the inter-chromosomal domain model should be renamed the inter-*chromatin* domain model, whereby chromosomes are invaginated with channels and subdivided into ~1 Mbp domains of chromatin [27].

1.3.4 Chromosome territory intermingling

A question that followed immediately on from the discovery of loops extending from chromosome territories was that of chromosome intermingling. Mathematical modelling approaches suggested that irradiation induced DNA damage corresponded to the intermingling of chromosome territories [62, 63]. This was later backed up with a study by Branco and Pombo, who studied thin cryosections of nuclei with high resolution light and electron microscopy; they found that there is significant intermingling of chromatin between chromosome territories, and that the extent of this intermingling correlated strongly with previously recorded irradiation induced translocation frequencies [64].

In support of chromosome intermingling, the *HoxB* extra-chromosomal loops found to extend from the chromosome territory upon gene activation [58] were found to make increased *trans* chromosomal interactions whilst looping out [65], suggesting that the loops contact other chromosomes rather than occupying an empty inter-chromosomal space.

1.4 Nuclear compartmentalisation

- Cajal bodies
- Speckles
- Nucleoli
- Polycomb / silencing bodies

1.4.1 Transcription factories

One nuclear subcompartment which has come to light within the past thirty years is the *transcription factory*, foci of hyper-phosphorylated RNA Polymerase II spread throughout the nucleus. The majority of geneic transcription appears to take place at transcription factories [66, 61, 67], challenging the classical model of transcription found in many text books.

The first study to suggest that Eukaryotic transcription does not take place with a processive RNA Polymerase moving along a DNA template was by Jackson, McCreedy and Cook in 1981. They showed that nascent RNA transcripts labelled with [^{3}H] uridine remained within the nucleus when loops of DNA were removed using a nuclease [69]. They went on to show that RNA Polymerase II and active genes were also resistant to elution after chromatin digestion [70]. The term 'transcription factories' was coined by Jackson *et al.* in 1993. Fluorescence microscopy was used to label the incorporation of bromouridine triphosphate (BrUTP) into nascent RNA; discrete foci of nascent transcription could be seen within the nucleus which did not form in the presence of the RNA Polymerase II inhibitor α-amanitin [66]. Further studies showed that these foci contain RNA Polymerase II along with many other components required for transcription [71, 72]. An ultrastructural study by Eskiw *et al.* used correlative microscopy with both electron spectroscopic imaging (ESI) and fluorescence microscopy to study nuclei sections. ESI can distinguish nitrogen and phosphorous atoms without labelling, and fluorescence light microscopy can visualise transcription through labelling BrUTP in nascent transcripts. The authors found that nascent RNA is almost always

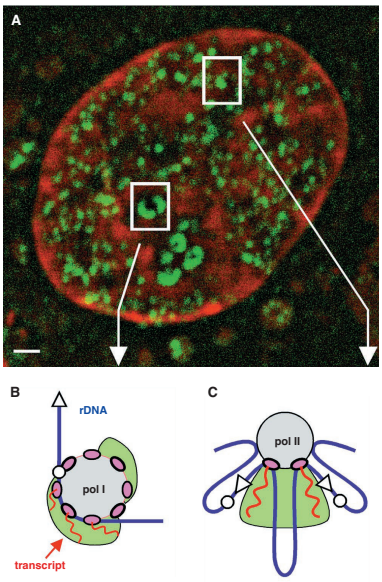

Figure 1.4 – Nascent RNA and transcription factories. (A) Transcription foci in HeLa cells, visualised with labelled Br-UTP in 100 nm cryosections. Nascent RNA (green) is concentrated in punctate foci. (B) Model for a nucleolar factory, showing a transcript with multiple polymerases generating a crescent shaped foci. (C) Model for a nucleoplasmic factory. Multiple transcribed regions each with a single polymerase generate a smaller cloud of nascent transcripts. Adapted from Cook *et al.*, Science (2009) [68].

associated with the surface of large nitrogen-rich protein structures with a diameter of ~87 nm, comparable in size to that predicted for a transcription factory [67].

The discovery of transcription factories has demanded a new model for the action of RNA Polymerase II [68]. The revised model proposes that instead of RNA Polymerase II freely diffusing to active genes and tracking along the gene body, genes are recruited to transcription factories and are pulled through a stationary polymerase. Such a model provides a better explanation for the mechanics of transcription - clustering of transcriptional activity may enable the cell to conduct transcription in a much more efficient manner; HeLa nuclei have a 1 μM concentration of active RNA Polymerase II, whereas the local concentration within transcription factories is closer to 1 mM [73]. Additionally, a polymerase enzyme moving along a gene would rotate with the helix of the DNA wrapping the nascent transcript around the template. Genes pulled through static transcription factories would extrude their RNA transcripts into the nucleoplasm [71, 68], creating topological loops within the template DNA which may be removed through the activity of topoisomerases.

Jackson *et al.* went on to work on a quantitative analysis of transcription factories in HeLa cells calculating the number of active RNA polymerases, the number of transcription sites and the number of polymerases associated with each transcriptional unit [74]. They showed that each HeLa cell nucleus contain ~ 2400 transcription factories, each with approximately 30 active RNA-Polymerase II complexes [74]. Importantly, this study showed that there are more transcribing units than there are foci of transcription, suggesting that genes must colocalise to transcribe. The number of transcription factories varies a great deal amongst cell types, but the observation that genes colocalise within transcription factories has been confirmed by a number of different techniques; Osborne *et al.* showed that genes situated within transcription factories are actively transcribed, whereas those outside are not [61]. They went on to demonstrate that transcription is a discontinuous process, with the frequency of nascent RNA transcription foci related to cellular mRNA concentrations, suggesting that transcription occurs in bursts. Multiple genes both in *cis* and in *trans* were seen to dynamically colocalise in transcription factories, supporting predictions that genes must share transcription factories [61, 74]. A later paper by Osborne *et al.* showed that the immediate-early gene *Myc* and *Fos* are dynamically recruited to existing transcription factories within 5 minutes of B-cell stimulation, suggesting that the recruitment of genes to pre-existing transcription factories may be a method of transcriptional control [75].

1.4.1.1 Specialised transcription factories

After the demonstration that genes share transcription factories, a number of groups postulated that specific genes may colocalise at a subset of transcription factories. The self-organising principle applied to transcription factories would predict this; just as the local concentration of active RNA Polymerase II is elevated by the clustering of transcribing units, the local concentration of transcription factors that bind to those transcribed units will also be elevated [76].

Genes that diffuse to a transcription factory already engaged with other genes that share the same factors are more likely to engage and be transcribed themselves [77]. Indeed, genes regulated by common transcription factors appear to cluster within specific chromosomes in yeast [22] and testis-specific genes are found clustered in *Drosophila* [23] - genes clustered in linear sequence are more likely to associate in three dimensions.

Some evidence that such specialised transcription factories may exist within mammalian cells came from Osborne *et al.* [75]. They showed that approximately one quarter of actively transcribing *Myc* alleles shared a transcription factory with *Igh*, over double the rate of colocalisation found with the control genes tested. Using DNA-FISH they showed an overall reduction in spacing between *Myc* and *Igh* alleles upon B cell stimulation, suggesting that *Myc* alleles are being specifically recruited to transcription factories containing transcribing *Igh* [75]. Xu and Cook later demonstrated *in-vitro* that cells transfected with thousands of plasmids containing different transcriptional units clustered together at a handful transcription factories, segregated according to their promoter and the presence of an intron [78].

Perhaps the best evidence for the existence of specialised transcription factories came in a publication by Schoenfelder *et al.* in 2010 [79]. They used a variety of techniques to investigate the nuclear localisation of erythroid genes. A genome-wide screen of genes associating with *Hba* and *Hbb* at transcription factories showed enrichment for genes with CACC motifs capable of binding the erythroid-specific transcription factor Klf1. Using immunofluorescence they found that nuclear Klf1 foci overlap with active RNA Polymerase II foci, suggesting that a subset of transcription factories are enriched for Klf1 in erythroid tissues (Fig 1.5). Co-localising Klf1 dependent genes associated with these Klf1 specific transcription factories at a high frequency, and a number of these gene associations were lost upon the knockout of Klf1 expression. These data suggest that a network of Klf1 specific transcription factories exist within mouse erythroid tissues, and that Klf1 specific genes are preferentially recruited to these sites [79]. If such specialised transcription factories are a general feature in mammalian nuclei, they could be a key driving force in the organisation of the genome.

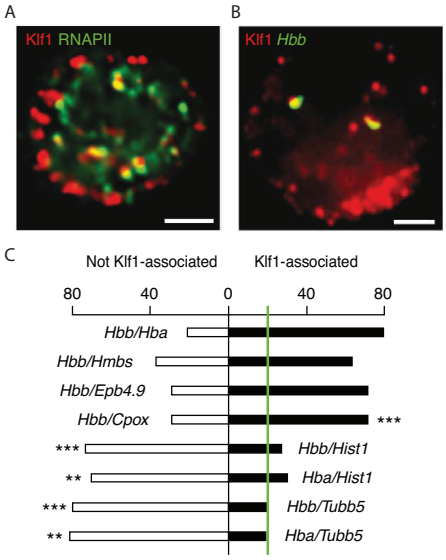

Figure 1.5 – Klf1 specialised transcription factories. (A) Immunofluorescence showing Klf1 (red) overlapping with a subset of RNA Polymerase II transcription factories (green). (B) RNA immuno-FISH showing transcribing *Hbb* alleles (green) co-localising with Klf1 (red). (C) Percentages of co-localising transcript pairs colocalisation with Klf1 associated transcription factories (right, black bars) and non-Klf1 associated factories (left, white bars). Expected level of association due to chance shown by green bar. Scale bars show in (A) and (B) represents 2 μm. Taken from Schoenfelder *et al.* [79].

It is important to note that the evidence above does not describe *exclusively* specific transcription factories, factories which are incapable of transcribing genes not controlled by

Klf1. Indeed, although statistically significant, the gene associations detected suggest a population bias. Specialised transcription factories are certainly not required for the transcription of the eukaryotic genome, but may aid efficiency and affect nuclear organisation and function.

1.5 Chromatin interactions in three dimensions

1.5.1 Promoter and enhancer interactions

The classical model of gene promoter regulation began with the characterisation of the *Lac* operon in *Escherichia coli* in 1961 [80]. Three genes required for the digestion of lactose are controlled by the binding of a repressor protein which blocks transcription by RNA Polymerase II. This repressor can form a tetramer and bind two region simultaneously, requiring a topological loop in the chromatin [81].

A number of Eukaryotic genes were later found to require distal enhancer elements [82] prompting speculation that the looping of chromatin allowing direct interaction between sequences may be a common mechanism [83]. Direct evidence for *in-vivo* chromatin interactions was provided in 2002 with two papers studying the mouse *Hbb* locus [84, 85]. The locus contains five beta-like genes β, γG, γA, δ and ε, which are arranged in order of their expression through development. Approximately 50 kilobases upstream is a locus control region (LCR), containing multiple DNase hypersensitive sites (HS1-6) [86]. Carter *et al.* used a novel technique called RNA TRAP (tagging and recovery of associated proteins) which localises horseradish peroxidase to the site of nascent RNA production of a specific gene. The horseradish peroxidase catalyses biotinylation of nearby chromatin which can be purified and quantified by quantitative real-time PCR (qPCR). They showed a 15-fold enrichment of biotin over the HS2 region of the LCR when the actively transcribing *Hbb-b1* transcript was used for the horseradish peroxidase localisation, demonstrating that this region of the LCR is in very close proximity to the *Hbb-b1* gene [84].

Tolhuis *et al.* used the chromatin-conformation-capture technique (3C) developed by Dekker *et al.* to study the organisation of chromatin within yeast [87] (for discussion of this technique see Section **??**). They studied the same globin locus and also found evidence for interaction between the LCR hypersensitive regions and the *Hbb* genes. Importantly they showed that in mouse foetal brain tissue, where the *Hbb* locus is not expressed, the chromatin adopted a linear conformation without any looping [85]. Palstra *et al.* went on to show that each gene contacts the LCR as it is expressed, supporting a system where the developmentally controlled genes must contact an enhancer within the LCR in order to be expressed [88]. They showed that erythroid progenitor cells dedicated to the lineage but not yet expressing beta-like globin genes form a 'poised' structure, contacting the LCR but not engaging strongly with the HS2 enhancer [88].

1.5.2 Chromatin hubs

The beta globin locus has become a model system for the active chromatin hub (ACH) model [89], describing a system where loops of chromatin containing elements capable of controlling the expression of genes are held in close three-dimensional space to the genes that they control. Other gene clusters have been shown to behave in a similar manner, notably the T_H2 locus and *Hox* clusters.

The T_H2 locus control region is involved in the transcriptional control of cytokine genes IL4, 5 and 13 [90]. Spilianakis *et al.* showed that these genes co-associate in T cells, NK cells, B cells and fibroblasts, despite not being expressed in B cells and fibroblasts. In T cells and NK cells expressing these genes, they associate with the T_H2 LCR [91]. The authors suggest that this mechanism allows the coordinate expression of the gene cluster in a controlled manner.

The *Hox* genes are master regulators of gene transcription and are responsible for the creation of vertebrate segments during development. *Hox* gene clusters *A* to *D* are transcribed in sequence as development of the embyro progresses and have been observed decondensing and looping out from their chromosome territory upon activation [58]. Noordermeer *et al.* studied the *Hox* clusters in three different mouse embryonic day 10.5 tissues; forebrain, anterior trunk and posterior trunk [92]. They found that the *Hoxd* cluster formed a discrete domain in forebrain, where it is inactive. In anterior and posterior trunk the *Hoxd* cluster is active, but different genes are transcribed. In both tissues they found the cluster to form two distinct compartments correlating with an inactive and active regions. Using 4C, a technique based on chromosome conformation capture (3C), they showed that genes move from the inactive to active compartment as they are activated, correlating with active histone marks [92].

1.5.3 Interactions in *trans*

As the role for distal enhancers has become more established, evidence has been uncovered for increasingly distant interactions. Lettice *et al.* used data from known chromosomal translocations resulting in polydactyly to find and characterise an enhancer element for the *SHH* gene over a megabase upstream, within an intron of another gene [93]. Spilianakis *et al.* went on from characterising the intra-chromosomal interactions of the T_H2 LCR to show that the same locus forms inter-chromosomal interactions [94]. Depending on the stimulus received, naïve T cells can differentiate into either TH1 or TH2 cells, defined by the expression of either IFN-γ or IL-4. Spiliankis *et al.* showed that the *Ifng* gene on chromosome 10 can interact with the T_H2 LCR on chromosome 11 to stimulate *Ifng* expression whilst inhibiting *IL4* expression. This interaction is the first interchromosomal interaction known to regulate gene expression [94]. A number of studies have since shown specific interchromosomal contacts involved in a number of processes ranging from X-inactivation to genomic imprinting, showing that these interactions play an important role in chromatin biology [95].

1.5.4 Global interaction maps

Our understanding of the three-dimensional organisation of the genome has advanced in leaps and bounds during the last decade largely because of the development of chromatin conformation capture (3C) and its derivatives [96]. There are a large number of 3C variants, but they can be grossly categorised into four classes based on how many loci can be interrogated in a single experiment: one-to-one (3C, quantitative 3C), one-to-all (4C, e4C, ACT), many-to-many (5C) and all-to-all (ChIA-PET, Hi-C, TCC).

The recent development of all-to-all methods has allowed the confirmation of the entire genome to be probed in a single experiment. This approach has many advantages; such an unbiased approach allows the detection of unexpected interactions and associations can be probed in parallel allowing a far higher rate of data collection. Whilst these techniques have had great impact on the field, they are currently limited by the depth of sequencing that is achievable with today's technology. To address this Sanyal *et al.* recently published a paper as part of the ENCODE project describing the interaction profiles of 628 transcription start sites (TSS) and 4535 surrounding fragments, representing approximately 1% of the genome [97]. To achieve the resolution required for the robust detection of promoter-element interactions, Sanyal *et al.* used 5C, a many-to-many technique that uses a library of oligonucleotides with common adapters to anneal to 3C products and create a library capable of being sequenced. They sequenced libraries from three ENCODE cell lines: K562, HeLa-S3 and GM12878. Only a small proportion of the looping interactions uncovered were shared between the three cell types, with ~60% of interactions being unique to a single cell line. The majority of TSS looping interactions could be classified as interacting with enhancer elements, promoters or regions bound by the structural protein CTCF. Looping interactions with enhancer elements were significantly enriched for actively expressed TSS, demonstrating the importance of three-dimensional chromatin contacts in the regulation of gene expression.

1.6 What drives nuclear organisation?

As our understanding of the structure of the nucleus evolves, an increasing number of structural features and patterns are uncovered. Teasing apart correlation and causation to find the driving forces behind nuclear organisation is not an easy task and remains a hot topic within the field.

1.6.1 Transcription

The discovery of transcription factories has changed our view of nuclear organisation substantially. If genetic templates are mobile and transcription factories are fixed, then the nucleus may be able to use the process of transcription as a tool to fold the genome into specific conformations.

Kimura *et al.* quantified the amount of stable RNA Polymerase II in HeLa cells [98] adding to the work by Jackson *et al.* demonstrating the stability of nascent transcripts and Polymerase in the nucleus [69, 70]. Mitchell and Fraser later demonstrated that RNA Polymerase II transcription factories remain in the absence of transcription, though gene association with factories is ablated if transcription initiation is inhibited [99]. These data, along with the observation that genes are recruited to pre-existing transcription factories upon activation [61], support a model whereby transcription factories are attached to a relatively immobile nuclear substructure. This means that RNA Polymerase II can act as a motor, dragging template chromatin through the nucleus as it is transcribed, powered by the removal of phosphate groups during RNA synthesis [68]. Yin *et al.* measured the force produced by a single *E. coli* RNA Polymerase using an immobilised enzyme transcribing a template bound to a polystyrene bead held by optical tweezers [100]. RNA Polymerase stalled when the force applied was greater than 14 pN, making RNA Polymerase the most powerful biological motor known. Papantonis *et al.* demonstrated the potential of RNA Polymerase to pull chromatin transcripts through the nucleus *in vivo* by using 3C to measure the change in association between regions of DNA after activation of the TNFα gene [101]. They found that as the gene was transcribed, downstream regions of chromatin progressively came into contact with other transcribing regions at the transcription factory.

1.6.1.1 Not finished - write about transcription stop cassette in T cells

A good example of transcription actively organising a structural motif is within the immunoglobulin heavy-chain locus. The *Igh* locus contains variable (V_H), diversity (D_H), joining (J_H) and constant (C_H) coding elements which must contact each other for the process of recombination. The *Igh* locus spans 3 Mbp and forms a rosette-like structure [102]. This structure appears to be driven by association at a transcription factory, with low level antisense transcription of the V_H gene segments prior to recombination [#REF]. Interestingly, the level of antisense transcription correlates with the distance of the V_H gene from the $C_H$region, and if this transcription is inhibited, the $V_H$gene usage is biased towards V_H genes closer to the C_H region in the linear DNA template [#REF]. This data suggests a model whereby the antisense transcription of the V_H genes is required to bring them into contact with the C_H region to give equal probability of recombination regardless of the linear spacing of the genes. [1]

The role for transcription in the organisation of the genome is further supported by the pervasive nature of transcription. As much as 15% of the human genome is actively transcribed [103], including many enhancers [104, 105]. In the recent ENCODE study of TSS interactions, Sanyal *et al.* found that enhancer elements looping to a TSS were significantly more likely to express enhancer RNAs [97], supporting a model that the transcription of a chromatin template could be responsible for the formation of chromatin loops.

[1] This will need to be re-written after I've found the right references.

An attractive model for larger scale genome organisation revolves around the transcription of housekeeping and tissue-specific genes. Lercher *et al.* have shown that genes with high expression in multiple tissue types have a propensity be present in clusters within the genome [106] and the mouse alpha globin locus has been shown to assemble at a transcription factory already transcribing a cluster of housekeeping genes in erythroid cells [107]. These housekeeping genes were bound stably to the nuclear matrix, whereas the alpha globin genes were eluted in a high salt extraction [108]. These studies may point to a model of the nucleus where persistently transcribed housekeeping genes are responsible for forming transcription factories, with tissue specific genes being recruited to these sites upon activation [108]. Such a system could have far reaching consequences, with chromosome conformation being determined by housekeeping gene hubs.

1.6.2 CTCF

- Insulators. Chicken beta globin.
- ChIA-PET - [109]
- CTCF not always marking insulators - interactions skipping CTCF sites in Sanyal *et al.*[97]
- Talk about CTCF bordering domains, Hi-C papers

1.6.3 Cohesin

- Introduction to standard stuff about sister chromatid adhesion
- Basically all of Suzie's papers showing that Cohesin binds CTCF in interphase

1.6.4 Actin and myosin

- Nuclear actin and myosin and stuff.

1.6.5 Replication

- Should probably write quite a bit here as haven't talked about it at all yet
- Similarity to transcription / factories etc

1.6.6 The self-organisation model

- Cook paper on entropy basics, nature of everything wanting to cluster. [76, 71]
- Transcription factor networks (Bertie).

- Mention specialised transcription factories again?

1.7 The annotated genome

- Talk about how wonderful systems biology is
- Integration of datasets - ChIP Seq is two dimensional and needs interaction data to be understood
- Correlation between different marks gives the full picture
- Genome wide studies paving the way for a different type of analysis
- Talk about the systems biology of networks
  - This is where I extoll the virtues of Bertie's work looking at transcription factor networks
  - Use *cis* regulatory modules as an example? Then shoot down with Dekker 5C paper? Check Bertie doesn't love them first (I think he might)

1.7.1 The ENCODE project

- Description of ENCODE project
- Couple of key findings
- Mention value of ENCODE datasets beyond initial publications

1.8 Chromosomal translocations

1.8.1 Formation of chromosomal translocations

Since the development of chromosome banding microscopy, it has been known that the genetic material within cancerous cells is frequently disrupted [110]. Chromosomal translocations involve the rearrangement of genetic material between non-homologous chromosomes through the formation and aberrant repair of two double strand breaks (DSBs). The resulting product can be a straight swap without any loss or gain of sequence (balanced translocation) or can result in deletions or gain of material (unbalanced translocation).

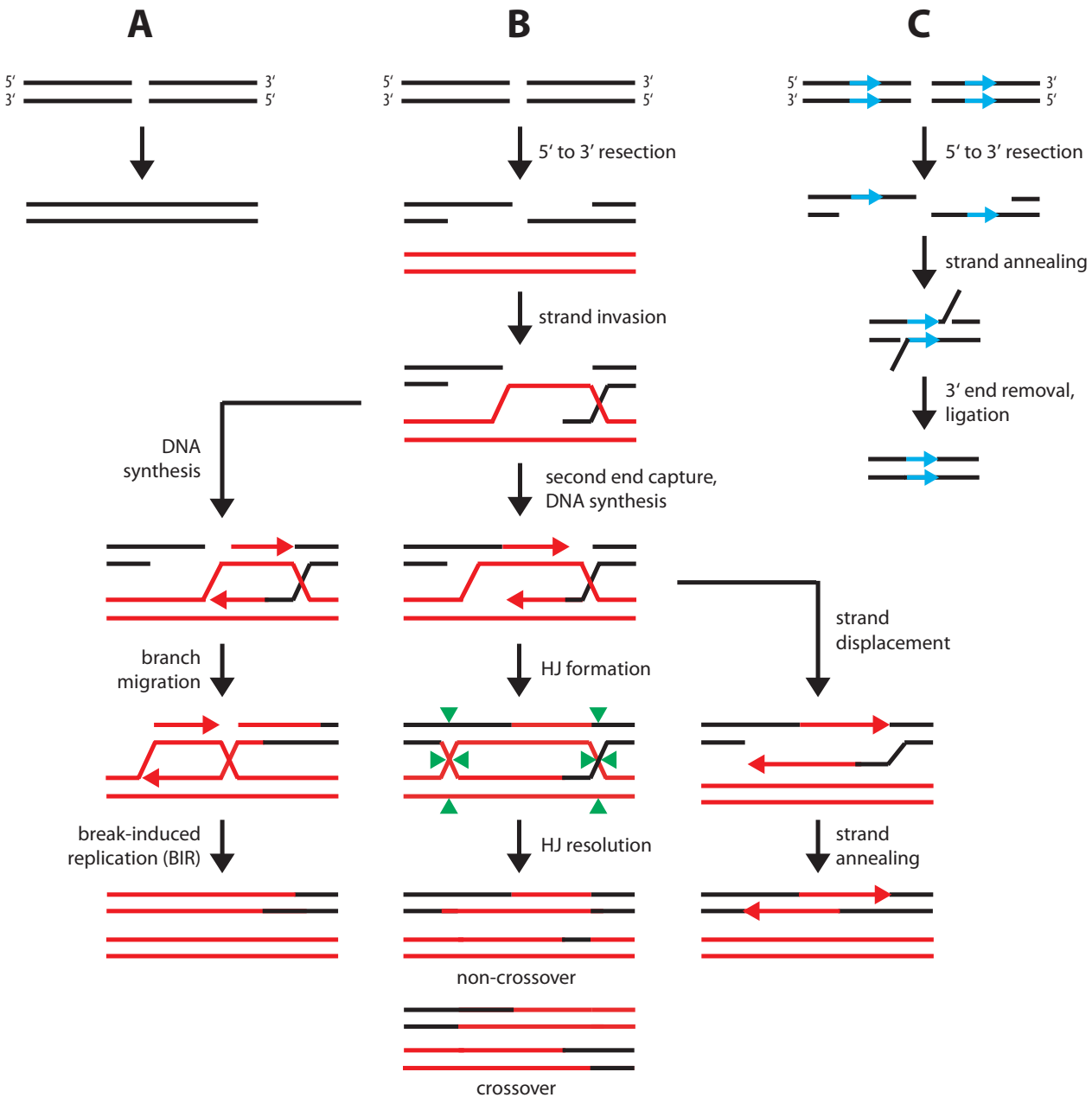

Figure 1.6 – **DSB repair.** Three double strand break repair pathways. **(A)** NHEJ (Non-Homologous End Joining) requires little or no sequence homology and joins any two nearby ends. **(B)** HR (Homologous Recombination) uses a strand of homologous DNA on a duplicated interphase chromosome or sister chromatid to reconstitute sequence before joining. Although it only requires a single double strand break to generate a crossover, the extensive region of homology needed makes errors rare. **(C)** SSA/NHEJ (Single Strand Annealing / Non-Homologous End Joining) searches for regions of homology, such as repeat sequences, and deletes any sequence between. From Longhese *et al.*, 2006 [112].

1.8.1.1 DSB repair

Double strand breaks can be caused by exogenous damage (ionizing radiation, free radicals) and endogenous damage (physiological programmed DSBs). As many as 1 million DNA lesions are formed per cell per day [111] and can be

highly deleterious to the cell, casuing loss of genetic material, translocations and ultimately cell death if left unrepaired. Mammalian cells have an array of different DSB repair pathways that depend on the cellular context and type of lesion [112].

Homologous recombination (HR, Fig 1.6B) repairs double strand breaks by using long regions of homology on undamaged sister chromatids or homologous chromosomes. HR is primarily active during cell replication and rarely leads to serious chromosomal rearrangements or translocations, although 'crossing-over' can sometimes occur, whereby two sections of homologous chromosomes switch to their partner chromosomes (a process that is encouraged in meiosis with the formation of chiasmata).

Single strand annealing (SSA, Fig 1.6C) creates a short region of single stranded DNA at the site of the DSB which is used to search for regions of microhomology, usually repeat sequences in the same orientation. These regions anneal and any excess single stranded tails are removed to allow the nicks to be ligated. SSA is prone to introducing deletions into the genome, and can form chromosomal translocations if two sequences with similar repeat regions are nearby, and both suffer simultaneous DSBs.

Non-homologous end joining (NHEJ, Fig 1.6A) is a last resort for the cell, whereby any two adjacent DSBs are ligated, needing little or no sequence homology. This pathway is extremely prone to errors as it is capable of joining any two strands of DNA without knowledge of their identity.

1.8.2 Chronic myeloid leukaemia

Chronic myeloid leukaemia (CML) is one of the best studied cancers, despite having a relatively low occurance [113]. The term 'leukaemia' was based on CML patients in the mid 1800s [114], and it was the first cancer for which a direct causative genotype was found [115]. A number of standardised therapies exist for CML and up to 87% of patients achieve complete cytogenetic remission [113], making it one of the success stories of cancer research.

1.8.2.1 The Philadelphia chromosome

The genetic abnormality associated with CML was first identified in 1960 by Nowell *et al.* [115]. They noticed the presence of an abnormal chromosome and called it the Philadelphia chromosome, or Ph chromosome. The chromosome was initially thought to be the result of a deletion until chromosome banding enabled its identification as a product of a translocation [116]. The breakpoints within chromosomes 9 and 22 were subsequently identified as being within the *c-ABL* and *BCR* genes, respectively. The *c-ABL1* gene (also known as *ABL1)* was named as such because of its similarity to the Abelson murine leukaemia virus gene *v-ABL [117].* Less is known about the *BCR* gene, so named because it was found at the Breakpoint Cluster Region of chromosome 22.

1.8.2.2 *BCR* and *ABL1*

The *ABL1* gene encodes a tyrosine kinase which is ubiquitously expressed in mammalian cells. Involved in cell cycle regulation, it is thought to be involved in several cell signalling pathways [118]. Less is known about the endogenous function of the *BCR* protein; it contains a serine-threonine kinase and has GTPase activity but its function is not known [118]. The t(9;22)(q34;q11) translocation creates a fusion protein lacking the SH2 domain of *ABL1* which normally regulates its activity. This results in constitutively active tyrosine-kinase activity which drives oncogenesis [118]. Relatively little is known about the mechanism of translocation formation, though ionising radiation is known to be a risk factor [119, 120].

Identification and characterisation of the *BCR-ABL* translocation and its fusion protein led to the development of a number of kinase inhibitors such as the drug *imatinib*, which achieve excellent success in the treatment of CML [113].

1.8.3 Mixed lineage leukaemia

Acute leukaemia is typically categorised as either Acute Lymphoblastic Leukaemia (ALL) or Acute Myeloid Leukaemia (AML) according to which lineage of blood cells are cancerous. ALL is defined by the uncontrolled proliferation of lymphoblasts, precursors to lymphocytes which differentiate into B cells, T cells and NK cells. ALL was estimated to account for approximately 12% of all leukaemia cases in the US in 2008 with around five and a half thousand cases occurring annually. It is the most common form of cancer in children aged under fourteen [121].

AML is a cancer of the myeloid lineages of blood cells, responsible for approximately 30% of all cases of leukaemia in the US for 2008 with a higher mortality rate than ALL (66% and 27% mortality for AML and ALL, respectively) [121]. The incidence of AML increases with age; the median age of AML diagnosis between 2002 and 2006 in the US was 67 years, and the median age of mortality was 72 [121].

The biology of leukaemia is not always as discrete as this classification however, some patients present with expansion of both lymphoid and myeloid lineages [122]. Common genetic abnormalities found in these patients include the t(9;22)(q34;q11) Philadelphia chromosome and structural changes in 11q23 [122]. Chromosomal translocations within band q23 of chromosome 11 have been implicated in both AML and ALL, all involve the gene *MLL* (also known as *ALL-1*, *Htrx*, *HRX*) identified by Ziemin-van der Poel *et al.* in 1991 [123]. Translocations involving the *MLL* gene are found in over 70% of all infant leukaemias [124] and approximately 10% of adult AML cases [125]. The translocation correlates with poor patient prognosis and is of high clinical interest [126].

In 2002, Armstrong *et al.* showed that acute lymphoblastic leukaemias containing a translocation within the *MLL* gene have a unique expression profile that is different to ALL and AML, and suggests an origin within a less committed progenitor cell which can produce cells in both the myeloid and lymphoid cell lineages [127]. They suggest that these leukaemias are substantially different from AML and ALL and deserve a new, distinct, class of leukaemia called Mixed

Lineage Leukaemia (MLL). Further cytogenetic studies have supported this theory; leukaemic cells in MLL have been found to express cell surface antigens normally present on both myeloid and lymphoid cells such as CD14 and CD19 [125].

The suggestion that MLL is initiated within a progenitor cell is in line with previous evidence showing that some cases of ALL and AML are initiated within undifferentiated haematopoietic stem cells (HSCs) [128, 129]. HSCs are present within the $CD34^+$ progenitor cell population, accounting for approximately 3% of normal human bone marrow and 0.3 to 0.5% of human cord blood mononuclear cells [130]. They are long lived and capable of self-renewal, differentiating into lineage restricted progenitors and eventually mature terminally differentiated white blood cells. HSCs are necessary for the long term maintenance of the haematopoietic system and are commonly used to repopulate bone marrow after myeloablative therapy, as well as in the treatment of a number of other disorders such as autoimmune, cardiac and vascular diseases [131]. Cancer stem cells are thought to be present in both leukaemias and solid tumours [132] and HSCs are a probable founder population due to their ability to self-renewal [133]. The concept of a small pool of cancer stem-cells driving the large heterogeneous pool of cancer cells has a number of implications for treatment - these are the cells that must be targeted for the efficient and long lasting cure of cancer [133].

1.8.3.1 The MLL protein

The MLL protein is a H3K4 methyltransferase involved in the positive regulation of global gene regulation, including the maintenance of expression of the *Hox* genes [134]. *MLL* is required for embryonic haematopoiesis [135] and adult bone marrow maintenance [136]. MLL is a mammalian homologue of the *Drosophila melanogaster* trithorax complex and is thought to bind DNA via an AT-hook domain [137] and zinc finger domain [138]. It's thought that the zinc finger domain targets MLL to unmethylated CpG islands DNA [138] and ChIP studies have shown that MLL binds to a subset of transcribed genes [139]. MLL binds promoters and gene bodies, associating tightly with RNA Polymerase II [139].

All known MLL fusion proteins contain exons 8-13 of *MLL* and in-frame exons of a partner gene [125]. Fusion proteins always retain their AT-hook and zinc-finger CxxC motifs, which are essential for their transforming potential [140]. The H3K4 methyltransferase domain of MLL is often lost in fusion proteins [125], despite this the fusion proteins can drive constitutive expression of *HOXA9* and *MEIS1*, which if over-expressed together in the absence of a MLL fusion protein give a similar phenotype [141].

1.8.3.2 The *MLL* gene

Translocations within the *MLL* gene are found within approximately 10% of all human leukaemias [142]. There are 87 documented *MLL* translocation partners of which 51 have been characterised at the molecular level [143]; the five most frequent translocation partners, *AF4*, *AF9*, *ENL*, *AF10* and *AF6* account for approximately 80% of cases [143].

Translocations within the *MLL* gene usually occur within an 8.3 kb *BamHI* fragment known as the breakpoint cluster region [144]. This region contains exons 5-11 as well as a number of repeat regions, notably 8 direct *Alu* SINE repeats, 5 direct L1 and L2 LINE repeats and 2 MER elements, as well as a number of putative Topoisomerase II binding sites and a SAR/MAR [145] (Fig 1.7). An internal promoter is present within the murine *Mll* breakpoint cluster region, correlating with etoposide-induced DSBs [146], DNase I hypersensitive sites [147] and histone modifications associated with transcription [148].

1.8.3.3 Mechanisms of *MLL* translocation formation

A clue to how translocations form within the *MLL* gene comes from the observation that 11q23 translocations are especially prevalent in therapy-related leukaemias - secondary leukaemias that develop in patients after treatment for a primary cancer with topoisomerase II inhibitors [125]. Topoisomerases are found in all eukaryotic nuclei and are able to relieve supercoiling and promote chromosome disentanglement [149]. They function by binding DNA, forming a transient double strand break and passing another strand of DNA through the gap, before ligating the double strand break.

Topoisomerases are important in transcription, which creates supercoils as template DNA is processed through static RNA Polymerase II enzymes [150]. DNA topoisomerase II associates with gene promoters [151] and is required for the transcription of genes longer than 3 Kbp in yeast [152]. Topoisomerase II induced double strand breaks have been implicated in the regulation of certain genes via the assembly of transcription complexes and changes in chromatin structure [153].

DNA topoisomerase II inhibitors are commonly used as chemotherapeutic agents and work by decreasing the ligation rate, disrupting the cleavage / ligation equilibrium. This leads to an increase in DNA cleavage and an accumulation of double strand breaks, triggering the cell DNA damage response and leading to cell death by apoptosis [154]. Topoisomerase II inhibitors are widely used chemotherapeutic agents, effective against a range of malignancies including small-cell lung cancer and gonadal tumours [155].

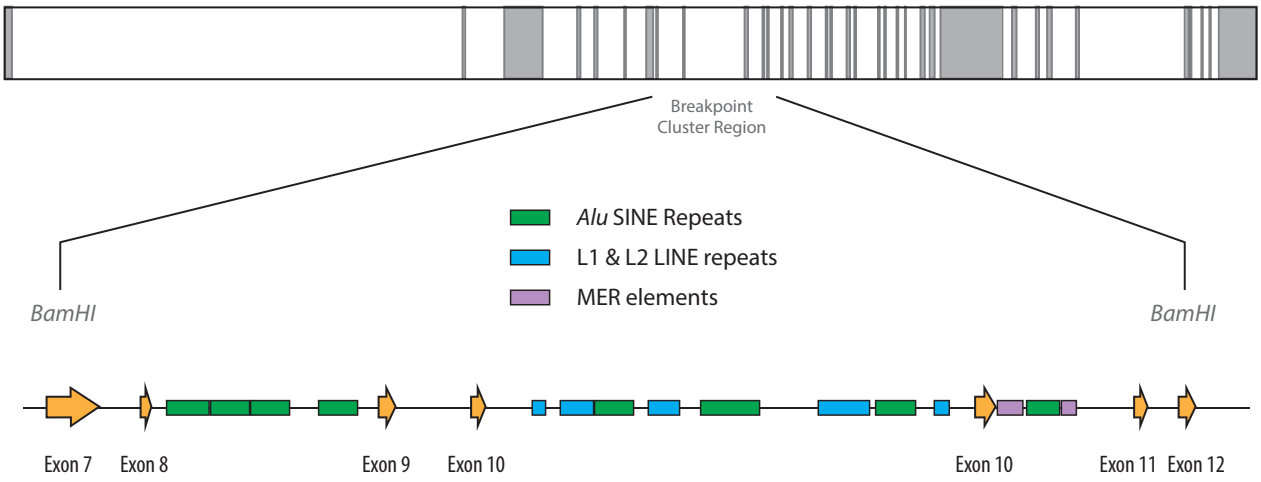

Figure 1.7 – Diagram of the *MLL* breakpoint cluster region.

The *MLL* breakpoint cluster region contains a number of putative topoisomerase II binding sites, prompting a number of groups to hypothesise a mechanism of topoisomerase inhibitor related translocations. Libura *et al.* showed that $CD34^{+}$ HSPCs exposed to etoposide formed 11q23 chromosomal aberrations typical of those seen in clinical samples [156]. They went on to show that such exposure to etoposide increased the proliferative potential of the cells in a bone marrow graft assay using immune-deficient mice [130]. This data supports a model whereby poisoned topoisomerase II may create double strand breaks within the *MLL* gene, allowing the NHEJ machinery to aberrantly repair the locus due to microhomology found in the nearby repetitive elements (Fig 1.7).

1.8.4 Effect of nuclear organisation on translocation formation

1.8.4.1 Breakage first and contact first models

For a chromosomal translocation to form, two double strand breaks must exist simultaneously and be adjacent in three-dimensional space. Two models have been described to explain how DSBs may meet in the nucleus – the breakage first model and the contact first model. The breakage first model states that double strand breaks can form anywhere and are able to freely diffuse in the nuclear space. They undergo large scale movement through the nucleus until they meet and are joined. The contact first model states that the two double strand breaks form in sequences already close to each other within the nuclear space. Large scale chromatin movements are not required for the two DSBs to meet.

In support of the breakage first model Aten *et al.* showed that a linear series of DSBs formed simultaneously in HeLa nuclei clustered [157], however, others have argued that this may be the result of a higher degree of chromatin mobility along the ion bean trajectory used to generate the DSBs [158]. A larger body of evidence supports a contact first model; DSBs created with ultra-soft X-rays by Nelms *et al.* remained in a fixed position for several hours after the damage was caused [159] and Jakob *et al.* used live cell microscopy to visualise proteins involved in DNA damage

signalling and repair. After accumulation at sites of DNA damage caused by heavy ion impacts the foci exhibited only a small degree of movement [158]. Soutoglou *et al.* showed a high degree of positional stability of DSB ends created by endonuclease digestion, with a greater degree of local diffusion seen in the absence of the Ku80 DNA-end binding protein [160].

1.8.4.2 Chromosome Territories

The large scale organisation of chromosome territories has been implicated in translocation frequency by a number of studies. Kozubek *et al.* showed that chromosomes 9 and 22 were found in the centre nuclei more frequently that would be expected by chance in lymphocytes, T- and B-cells, HL60 cells and bone marrow cells [161]. They used neutron irradiation to show that transfer of genetic material was much higher than that found with chromosome 8, which was found towards the nuclear periphery [161]. Parada *et al.* investigated the positions of chromosomes 12, 14 and 15 in a mouse lymphoma cell line and mouse splenocytes [162]. The found that two translocated chromosomes preferentially paired together in the nucleus of the cell line as well as in normal cells not containing the translocation. [162]. Parada *et al.* went on in a further study to examine the positioning of a larger range of chromosomes in a number of different tissue types [41]. They found that chromosome pairing was tissue specific and correlated with the occurrence of tissue-specific translocation events [41].

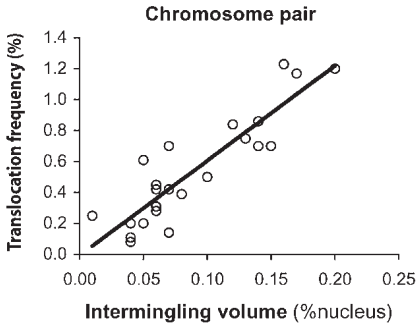

Figure 1.8 – Correlation of CT intermingling and radiation induced translocation frequencies. Adapted from Branco and Pombo, 2006 [64].

In 2006, Branco and Pombo published a study where they developed a new technique known as cryo-FISH to enhance the resolution of chromosome territory detection [64]. This technique had sufficient resolution to study the degree of intermingling between chromosome territories and they found that the degree of intermingling between chromosome pairs correlated strongly with their propensity to form translocations when subjected to radiation [64].

Interestingly, derivative chromosomes that result from balanced translocations affect the organisation of CTs within the nucleus [163], raising the possibility that the global changes in gene expression observed after oncogenic translocations could be in part due to changes in genome organisation [163].

1.8.4.3 Transcription Factories

Mounting evidence has shown that in addition to CT pairing, the position of specific genes involved in chromosomal translocations are frequently found in close proximity [164, 165]. As discussed in Section 1.6.1, localisation at transcription factories can drive organisation within the genome. Osborne *et al.* used DNA- and RNA-FISH to study the

localisation of the *Myc* proto-oncogene upon activation of B cells [75]. They showed that *Myc* on mouse chromosome 15 is dynamically recruited to an existing transcription factory and that this transcription factory was preferentially occupied by *Igh* on mouse chromosome 12. These observations of preferential association at transcription factories support to a model whereby genes sharing transcription factories are predisposed to the formation of chromosomal translocations [75]. Similar transcriptional associations have been observed for *MLL* and its frequent translocation partners *AF4* and *AF9* [166].

1.9 Hypothesis and aim of thesis

- State hypothesis - Preferential gene co-association within transcription factories predisposes specific gene pairs to the formation of chromosomal translocations.
- Aims - investigate the genome-wide association networks of specific oncogenes and see if gene pair co-localisation frequency matches up with gene-pair translocation frequency
  - Do this in an unbiased manner (not purely by RNA- and DNA-FISH), mention biased nature of such studies
  - Do this with a genome wide technique
  - Do this in the cell type where we think the translocations are taking place

Bibliography

[1] E. Heitz. Das heterochromatin der moose. *I Jahrb Wiss Botanik*, 69:762–818, 1928.

[2] L E Finlan, D Sproul, I Thomson, S Boyle, E Kerr, P Perry, B Ylstra, J R Chubb, and W A Bickmore. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet*, 4(3):e1000039, 2008.

[3] K L Reddy, J M Zullo, E Bertolino, and H Singh. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, 452(7184):243–7, March 2008.

[4] R Ileng Kumaran and David L Spector. A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *The Journal of cell biology*, 180(1):51–65, January 2008.

[5] C J Chesterton, B E Coupar, and P H Butterworth. Transcription of fractionated mammalian chromatin by mammalian ribonucleic acid polymerase. Demonstration of temperature-dependent rifampicin-resistant initiation sites in euchromatin deoxyribonucleic acid. *The Biochemical journal*, 143(1):73–81, October 1974.

[6] Curt a Davey, David F Sargent, Karolin Luger, Armin W Maeder, and Timothy J Richmond. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *Journal of molecular biology*, 319(5):1097–113, June 2002.

[7] a P Wolffe and J J Hayes. Chromatin disruption and modification. *Nucleic acids research*, 27(3):711–20, February 1999.

[8] B D Strahl and C D Allis. The language of covalent histone modifications. *Nature*, 403(6765):41–5, January 2000.

[9] Vicky W Zhou, Alon Goren, and Bradley E Bernstein. Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics*, 12(1):7–18, January 2011.

[10] Nathaniel D Heintzman, Rhona K Stuart, Gary Hon, Yutao Fu, Christina W Ching, R David Hawkins, Leah O Barrera, Sara Van Calcar, Chunxu Qu, Keith a Ching, Wei Wang, Zhiping Weng, Roland D Green,

Gregory E Crawford, and Bing Ren. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*, 39(3):311–8, March 2007.

[11] Jason Ernst and Manolis Kellis. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature biotechnology*, 28(8):817–25, August 2010.

[12] Gary Hon, Wei Wang, and Bing Ren. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS computational biology*, 5(11):e1000566, November 2009.

[13] Jason Ernst, Pouya Kheradpour, Tarjei S Mikkelsen, Noam Shoresh, Lucas D Ward, Charles B Epstein, Xiaolan Zhang, Li Wang, Robbyn Issner, Michael Coyne, Manching Ku, Timothy Durham, Manolis Kellis, and Bradley E Bernstein. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–9, May 2011.

[14] Paul B Talbert and Steven Henikoff. Histone variants–ancient wrap artists of the epigenome. *Nature reviews. Molecular cell biology*, 11(4):264–75, April 2010.

[15] Elaine M Dunleavy, Danièle Roche, Hideaki Tagami, Nicolas Lacoste, Dominique Ray-Gallet, Yusuke Nakamura, Yataro Daigo, Yoshihiro Nakatani, and Geneviève Almouzni-Pettinotti. HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell*, 137(3):485–97, May 2009.

[16] Daniel R Foltz, Lars E T Jansen, Aaron O Bailey, John R Yates, Emily a Bassett, Stacey Wood, Ben E Black, and Don W Cleveland. Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell*, 137(3):472–84, May 2009.

[17] Brian E Schwartz and Kami Ahmad. Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes & development*, 19(7):804–14, April 2005.

[18] Jordanka Zlatanova and Amit Thakar. H2A.Z: view from the top. *Structure (London, England : 1993)*, 16(2):166–79, February 2008.

[19] Maryse Adam, François Robert, Marc Larochelle, and L Gaudreau. H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. *Molecular and cellular biology*, 21(18):6270–9, September 2001.

[20] T Caspersson, L Zech, and C Johansson. Analysis of human metaphase chromosome set by aid of DNA-binding fluorescent agents. *Experimental cell research*, 62(2):490–2, December 1970.

[21] Barbara J Trask. Human cytogenetics: 46 chromosomes, 46 years and counting. *Nature reviews. Genetics*, 3(10):769–78, October 2002.

[22] Sarath Chandra Janga, Julio Collado-Vides, and M Madan Babu. Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41):15761–6, October 2008.

[23] Alexander M Boutanaev, Alla I Kalmykova, Yuri Y Shevelyov, and Dmitry I Nurminsky. Large clusters of co-expressed genes in the Drosophila genome. *Nature*, 420(6916):666–9, December 2002.

[24] E S Lander, L M Linton, B Birren, C Nusbaum, M C Zody, J Baldwin, K Devon, K Dewar, M Doyle, W FitzHugh, R Funke, D Gage, K Harris, a Heaford, J Howland, L Kann, J Lehoczky, R LeVine, P McEwan, K McKernan, J Meldrim, J P Mesirov, C Miranda, W Morris, J Naylor, C Raymond, M Rosetti, R Santos, a Sheridan, C Sougnez, N Stange-Thomann, N Stojanovic, a Subramanian, D Wyman, J Rogers, J Sulston, R Ainscough, S Beck, D Bentley, J Burton, C Clee, N Carter, a Coulson, R Deadman, P Deloukas, a Dunham, I Dunham, R Durbin, L French, D Grafham, S Gregory, T Hubbard, S Humphray, a Hunt, M Jones, C Lloyd, a McMurray, L Matthews, S Mercer, S Milne, J C Mullikin, a Mungall, R Plumb, M Ross, R Shownkeen, S Sims, R H Waterston, R K Wilson, L W Hillier, J D McPherson, M a Marra, E R Mardis, L a Fulton, a T Chinwalla, K H Pepin, W R Gish, S L Chissoe, M C Wendl, K D Delehaunty, T L Miner, a Delehaunty, J B Kramer, L L Cook, R S Fulton, D L Johnson, P J Minx, S W Clifton, T Hawkins, E Branscomb, P Predki, P Richardson, S Wenning, T Slezak, N Doggett, J F Cheng, a Olsen, S Lucas, C Elkin, E Uberbacher, M Frazier, R a Gibbs, D M Muzny, S E Scherer, J B Bouck, E J Sodergren, K C Worley, C M Rives, J H Gorrell, M L Metzker, S L Naylor, R S Kucherlapati, D L Nelson, G M Weinstock, Y Sakaki, a Fujiyama, M Hattori, T Yada, a Toyoda, T Itoh, C Kawagoe, H Watanabe, Y Totoki, T Taylor, J Weissenbach, R Heilig, W Saurin, F Artiguenave, P Brottier, T Bruls, E Pelletier, C Robert, P Wincker, D R Smith, L Doucette-Stamm, M Rubenfield, K Weinstock, H M Lee, J Dubois, a Rosenthal, M Platzer, G Nyakatura, S Taudien, a Rump, H Yang, J Yu, J Wang, G Huang, J Gu, L Hood, L Rowen, a Madan, S Qin, R W Davis, N a Federspiel, a P Abola, M J Proctor, R M Myers, J Schmutz, M Dickson, J Grimwood, D R Cox, M V Olson, R Kaul, N Shimizu, K Kawasaki, S Minoshima, G a Evans, M Athanasiou, R Schultz, B a Roe, F Chen, H Pan, J Ramser, H Lehrach, R Reinhardt, W R McCombie, M de la Bastide, N Dedhia, H Blöcker, K Hornischer, G Nordsiek, R Agarwala, L Aravind, J a Bailey, a Bateman, S Batzoglou, E Birney, P Bork, D G Brown, C B Burge, L Cerutti, H C Chen, D Church, M Clamp, R R Copley, T Doerks, S R Eddy, E E Eichler, T S Furey, J Galagan, J G Gilbert, C Harmon, Y Hayashizaki, D Haussler, H Hermjakob, K Hokamp, W Jang, L S Johnson, T a Jones, S Kasif, a Kaspryzk, S Kennedy, W J Kent, P Kitts, E V Koonin, I Korf, D Kulp, D Lancet, T M Lowe, a McLysaght, T Mikkelsen, J V Moran, N Mulder, V J Pollara, C P Ponting, G Schuler, J Schultz, G Slater, a F Smit, E Stupka, J Szustakowski, D Thierry-Mieg, J Thierry-Mieg, L Wagner, J Wallis, R Wheeler, a Williams, Y I Wolf, K H Wolfe, S P Yang, R F Yeh, F Collins, M S Guyer, J Peterson, a Felsenfeld, K a Wetterstrand, a Patrinos, M J Morgan, P de Jong, J J Catanese,

K Osoegawa, H Shizuya, S Choi, Y J Chen, and J Szustakowki. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, February 2001.

[25] Rogier Versteeg, Barbera D C van Schaik, Marinus F van Batenburg, Marco Roos, Ramin Monajemi, Huib Caron, Harmen J Bussemaker, and Antoine H C van Kampen. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome research*, 13(9):1998–2004, September 2003.

[26] H Caron, B van Schaik, M van der Mee, F Baas, G Riggins, P van Sluis, M C Hermus, R van Asperen, K Boon, P a Voûte, S Heisterkamp, a van Kampen, and R Versteeg. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science (New York, N.Y.)*, 291(5507):1289–92, February 2001.

[27] T Cremer and C Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews. Genetics*, 2(4):292–301, April 2001.

[28] Carl Rabl. Uber Zelltheilung. *Morphologisches Jahrbuch*, 10:214–330, 1885.

[29] S M Stack, D B Brown, and W C Dewey. Visualization of interphase chromosomes. *Journal of cell science*, 26:281–99, August 1977.

[30] T Cremer, C Cremer, H Baumann, E K Luedtke, K Sperling, V Teuber, and C Zorn. Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Human genetics*, 60(1):46–56, January 1982.

[31] Margit Schardin, T Cremer, H D Hager, and M Lang. Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. *Human genetics*, 71(4):281–7, January 1985.

[32] L Manuelidis. Individual interphase chromosome domains revealed by in situ hybridization. *Hum Genet*, 71(4):288–293, 1985.

[33] Andreas Bolzer, Gregor Kreth, Irina Solovei, Daniela Koehler, Kaan Saracoglu, Christine Fauth, Stefan Müller, Roland Eils, Christoph Cremer, Michael R Speicher, and Thomas Cremer. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS biology*, 3(5):e157, May 2005.

[34] H B Sun, J Shen, and H Yokota. Size-dependent positioning of human chromosomes in interphase nuclei. *Biophysical Journal*, 79(1):184–190, July 2000.

[35] M Cremer, J von Hase, T Volm, A Brero, G Kreth, J Walter, C Fischer, I Solovei, C Cremer, and T Cremer. Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome research :*

an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology, 9(7):541–67, January 2001.

[36] J A Croft, J M Bridger, S Boyle, P Perry, P Teague, and W A Bickmore. Differences in the localization and morphology of chromosomes in the human nucleus. *Journal of Cell Biology*, 145(6):1119–1131, 1999.

[37] João Ferreira, Giovanni Paolella, Carlos Ramos, and Angus I Lamond. Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *The Journal of cell biology*, 139(7):1597–610, December 1997.

[38] a E Visser, R Eils, A Jauch, G Little, P J Bakker, T Cremer, and J a Aten. Spatial distributions of early and late replicating chromatin in interphase chromosome territories. *Experimental cell research*, 243(2):398–407, September 1998.

[39] H Tanabe, F A Habermann, I Solovei, M Cremer, and T Cremer. Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications. *Mutation Research*, 504(1-2):37–45, 2002.

[40] Masahiko Kuroda, Hideyuki Tanabe, Keiichi Yoshida, Kosuke Oikawa, Akira Saito, Tomoharu Kiyuna, Hiroshi Mizusawa, and Kiyoshi Mukai. Alteration of chromosome positioning during adipocyte differentiation. *Journal of cell science*, 117(Pt 24):5897–903, November 2004.

[41] Luis a Parada, Philip G McQueen, and Tom Misteli. Tissue-specific spatial organization of genomes. *Genome biology*, 5(7):R44, January 2004.

[42] S T Kosak, J A Skok, K L Medina, R Riblet, M M Le Beau, A G Fisher, and H Singh. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*, 296(5565):158–162, 2002.

[43] Steffen Dietzel, Kourosh Zolghadr, Claudia Hepperger, and Andrew S Belmont. Differential large-scale chromatin compaction and intranuclear positioning of transcribed versus non-transcribed transgene arrays containing beta-globin regulatory sequences. *Journal of cell science*, 117(Pt 19):4603–14, September 2004.

[44] Daniele Zink, Margarida D Amaral, Andreas Englmann, Susanne Lang, Luka a Clarke, Carsten Rudolph, Felix Alt, Kathrin Luther, Carla Braz, Nicolas Sadoni, Joseph Rosenecker, and Dirk Schindelhauer. Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *The Journal of cell biology*, 166(6):815–25, September 2004.

[45] Sonja Stadler, Verena Schnapp, Robert Mayer, Stefan Stein, Christoph Cremer, Constanze Bonifer, Thomas Cremer, and Steffen Dietzel. The architecture of chicken chromosome territories changes during differentiation. *BMC cell biology*, 5(1):44, November 2004.

[46] I Szczerbal, H A Foster, and J M Bridger. The spatial repositioning of adipogenesis genes is correlated with their expression status in a porcine mesenchymal stem cell adipogenesis model system. *Chromosoma*, 118(5):647–663, 2009.

[47] W F Marshall, A Straight, J F Marko, J Swedlow, A Dernburg, A Belmont, A W Murray, D A Agard, and J W Sedat. Interphase chromosomes undergo constrained diffusional motion in living cells. *Current biology : CB*, 7(12):930–9, December 1997.

[48] Jonathan R Chubb, Shelagh Boyle, Paul Perry, and Wendy a Bickmore. Chromatin motion is constrained by association with nuclear compartments in human cells. *Current biology : CB*, 12(6):439–45, March 2002.

[49] Julio Vazquez, Andrew S Belmont, and John W Sedat. Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Current biology : CB*, 11(16):1227–39, August 2001.

[50] Chien-Hui Chuang, Anne E Carpenter, Beata Fuchsova, Terezina Johnson, Primal de Lanerolle, and Andrew S Belmont. Long-range directional movement of an interphase chromosome site. *Current biology : CB*, 16(8):825–31, April 2006.

[51] I S Mehta, M Amira, A J Harvey, and J M Bridger. Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts. *Genome Biol*, 11(1):R5, 2010.

[52] R M Zirbel, U R Mathieu, A Kurz, T Cremer, and P Lichter. Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 1(2):93–106, July 1993.

[53] T. Cremer, A. Kurz, R. Zirbel, S. Dietzel, B. Rinke, E. Schrock, M.R. Speicher, U. Mathieu, A. Jauch, P. Emmerich, H. Scherthan, T. Ried, C. Cremer, and P. Lichter. Role of Chromosome Territories in the Functional Compartmentalization of the Cell Nucleus. *Cold Spring Harbor Symposia on Quantitative Biology*, 58(0):777–792, January 1993.

[54] Anette Kurz, Stefan Lampel, J E Nickolenko, Joachim Bradl, A Benner, R M Zirbel, T Cremer, and P Lichter. Active and inactive genes localize preferentially in the periphery of chromosome territories. *The Journal of cell biology*, 135(5):1195–205, December 1996.

[55] Joanna M Bridger, Harald Herrmann, Christian Münkel, and Peter Lichter. Identification of an interchromosomal compartment by polymerization of nuclear-targeted vimentin. *Journal of cell science*, 111 (Pt 9:1241–53, May 1998.

[56] Emanuela V Volpi, Edith Chevret, Tania Jones, Radost Vatcheva, Jill Williamson, Stephan Beck, R Duncan Campbell, Michelle Goldsworthy, Stephen H Powis, Jiannis Ragoussis, John Trowsdale, and D Sheer. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *Journal of cell science*, 113 (Pt 9:1565–76, May 2000.

[57] Nicola L Mahy, Paul E Perry, and Wendy a Bickmore. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *The Journal of cell biology*, 159(5):753–63, December 2002.

[58] Séverine Chambeyron and Wendy a Bickmore. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes & development*, 18(10):1119–30, May 2004.

[59] Pernette J Verschure, I van Der Kraan, E M Manders, and R van Driel. Spatial relationship between transcription sites and chromosome territories. *The Journal of cell biology*, 147(1):13–24, October 1999.

[60] Nicola L Mahy, Paul E Perry, Susan Gilchrist, Richard a Baldock, and Wendy a Bickmore. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *The Journal of cell biology*, 157(4):579–89, May 2002.

[61] Cameron S Osborne, Lyubomira Chakalova, Karen E Brown, David Carter, Alice Horton, Emmanuel Debrand, Beatriz Goyenechea, Jennifer a Mitchell, Susana Lopes, Wolf Reik, and Peter Fraser. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature genetics*, 36(10):1065–71, 2004.

[62] W R Holley, I S Mian, S J Park, B Rydberg, and A Chatterjee. A model for interphase chromosomes and evaluation of radiation-induced aberrations. *Radiation research*, 158(5):568–80, November 2002.

[63] Lynn Hlatky, Rainer K Sachs, Mariel Vazquez, and Michael N Cornforth. Radiation-induced chromosome aberrations: insights gained from biophysical modeling. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 24(8):714–23, August 2002.

[64] Miguel R Branco and Ana Pombo. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS biology*, 4(5):e138, May 2006.

[65] Hugo Würtele and Pierre Chartrand. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome research : an international*

journal on the molecular, supramolecular and evolutionary aspects of chromosome biology, 14(5):477–95, January 2006.

[66] DA Jackson, AB Hassan, RJ Errington, and PR Cook. Visualization of focal sites of transcription within human nuclei. *The EMBO Journal*, 12(3):1059, 1993.

[67] Christopher H Eskiw, Alexander Rapp, David R F Carter, and Peter R Cook. RNA polymerase II activity is located on the surface of protein-rich transcription factories. *Journal of cell science*, 121(Pt 12):1999–2007, 2008.

[68] P. R. Cook. The Organization of Replication and Transcription. *Science*, 284(5421):1790–1795, June 1999.

[69] D A Jackson, S J McCready, and P R Cook. RNA is synthesized at the nuclear cage. *Nature*, 292(5823):552–555, 1981.

[70] DA Jackson and PR Cook. Transcription occurs at a nucleoskeleton. *The EMBO Journal*, 4(4):919, 1985.

[71] F J Iborra, a Pombo, D a Jackson, and P R Cook. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *Journal of cell science*, 109 (Pt 6:1427–36, June 1996.

[72] M a Grande, I van der Kraan, L de Jong, and R van Driel. Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II. *Journal of cell science*, 110 (Pt 1:1781–91, August 1997.

[73] David R F Carter, Christopher Eskiw, and Peter R Cook. Transcription factories. *Biochemical Society transactions*, 36(Pt 4):585–9, August 2008.

[74] D A Jackson, F J Iborra, E M Manders, and P R Cook. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Molecular biology of the cell*, 9(6):1523–36, June 1998.

[75] Cameron S Osborne, Lyubomira Chakalova, Jennifer a Mitchell, Alice Horton, Andrew L Wood, Daniel J Bolland, Anne E Corcoran, and Peter Fraser. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS biology*, 5(8):e192, 2007.

[76] Peter R Cook. Predicting three-dimensional genome structure from transcriptional activity. *Nature genetics*, 32(3):347–52, November 2002.

[77] Jon Bartlett, Jelena Blagojevic, David Carter, Christopher Eskiw, Maud Fromaget, Christy Job, Monee Shamsher, Inês Faro Trindade, Meng Xu, and Peter R Cook. Specialized transcription factories. *Biochemical Society symposium*, 181(73):67–75, January 2006.

[78] Meng Xu and Peter R Cook. Similar active genes cluster in specialized transcription factories. *The Journal of cell biology*, 181(4):615–23, May 2008.

[79] Stefan Schoenfelder, Tom Sexton, Lyubomira Chakalova, Nathan F Cope, Alice Horton, Simon Andrews, Sreenivasulu Kurukuti, Jennifer a Mitchell, David Umlauf, Daniela S Dimitrova, Christopher H Eskiw, Yanquan Luo, Chia-Lin L Wei, Yijun Ruan, James J Bieker, and Peter Fraser. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*, 42(1):53–61, January 2010.

[80] François Jacob and Jacques Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3):318–356, June 1961.

[81] Michael a Savageau. Design of the lac gene circuit revisited. *Mathematical biosciences*, 231(1):19–38, May 2011.

[82] J Banerji, S Rusconi, and W Schaffner. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27(2 Pt 1):299–308, December 1981.

[83] D Dorsett. Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Current opinion in genetics & development*, 9(5):505–14, October 1999.

[84] David Carter, Lyubomira Chakalova, Cameron S Osborne, Yan-feng Dai, and Peter Fraser. Long-range chromatin regulatory interactions in vivo. *Nature genetics*, 32(4):623–6, 2002.

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

[86] M a Bender, M Bulger, J Close, and M Groudine. Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Molecular cell*, 5(2):387–93, February 2000.

[87] Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. Capturing chromosome conformation. *Science*, 295(5558):1306–1311, February 2002.

[88] Robert-Jan Palstra, Bas Tolhuis, Erik Splinter, Rian Nijmeijer, Frank Grosveld, and Wouter de Laat. The beta-globin nuclear compartment in development and erythroid differentiation. *Nature genetics*, 35(2):190–4, October 2003.

[89] Wouter de Laat and Frank Grosveld. Spatial organization of gene expression: the active chromatin hub. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 11(5):447–59, January 2003.

[90] Gap Ryol Lee, Patrick E Fields, Thomas J Griffin, and Richard a Flavell. Regulation of the Th2 cytokine locus by a locus control region. *Immunity*, 19(1):145–53, July 2003.

[91] Charalampos G Spilianakis and Richard a Flavell. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nature immunology*, 5(10):1017–27, October 2004.

[92] D. Noordermeer, M. Leleu, E. Splinter, J. Rougemont, W. De Laat, and D. Duboule. The Dynamic Architecture of Hox Gene Clusters. *Science*, 334(6053):222–225, October 2011.

[93] Laura A Lettice, Simon J H Heaney, Lorna A Purdie, Li Li, Philippe de Beer, Ben A Oostra, Debbie Goode, Greg Elgar, Robert E Hill, and Esther de Graaff. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human molecular genetics*, 12(14):1725–35, July 2003.

[94] Charalampos G Spilianakis, Maria D Lalioti, Terrence Town, Gap Ryol Lee, and Richard a Flavell. Interchromosomal associations between alternatively expressed loci. *Nature*, 435(7042):637–45, June 2005.

[95] Robert Schneider and Rudolf Grosschedl. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes & development*, 21(23):3027–43, December 2007.

[96] Cameron S Osborne, Philip A Ewels, and Alice N C Young. Meet the neighbours: tools to dissect nuclear structure and function. *Briefings in functional genomics*, 10(1):11–7, January 2011.

[97] Amartya Sanyal, Bryan R. Lajoie, Gaurav Jain, and Job Dekker. The long-range interaction landscape of gene promoters. *Nature*, 489(7414):109–113, September 2012.

[98] Hiroshi Kimura, Yong Tao, Robert G Roeder, and P R Cook. Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. *Molecular and cellular biology*, 19(8):5383–92, August 1999.

[99] Jennifer a Mitchell and Peter Fraser. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes & development*, 22(1):20–5, 2008.

[100] H Yin, M D Wang, K Svoboda, R Landick, S M Block, and J Gelles. Transcription against an applied force. *Science (New York, N.Y.)*, 270(5242):1653–7, December 1995.

[101] Argyris Papantonis, Joshua D Larkin, Youichiro Wada, Yoshihiro Ohta, Sigeo Ihara, Tatsuhiko Kodama, and Peter R Cook. Active RNA polymerases: mobile or immobile molecular machines? *PLoS biology*, 8(7):e1000419, January 2010.

[102] Suchit Jhunjhunwala, Menno C van Zelm, Mandy M Peak, Steve Cutchin, Roy Riblet, Jacques J M van Dongen, Frank G Grosveld, Tobias a Knoch, and Cornelis Murre. The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell*, 133(2):265–79, April 2008.

[103] Jill Cheng, Philipp Kapranov, Jorg Drenkow, Sujit Dike, Shane Brubaker, Sandeep Patel, Jeffrey Long, David Stern, Hari Tammana, Gregg Helt, Victor Sementchenko, Antonio Piccolboni, Stefan Bekiranov, Dione K Bailey, Madhavan Ganesh, Srinka Ghosh, Ian Bell, Daniela S Gerhard, and Thomas R Gingeras. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science (New York, N.Y.)*, 308(5725):1149–54, May 2005.

[104] Jianhua Ling, Boris Baibakov, Wenhu Pi, Beverly M Emerson, and Dorothy Tuan. The HS2 enhancer of the beta-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a cis-linked globin promoter. *Journal of molecular biology*, 350(5):883–96, July 2005.

[105] Tae-Kyung Kim, Martin Hemberg, Jesse M Gray, Allen M Costa, Daniel M Bear, Jing Wu, David a Harmin, Mike Laptewicz, Kellie Barbara-Haley, Scott Kuersten, Eirene Markenscoff-Papadimitriou, Dietmar Kuhl, Haruhiko Bito, Paul F Worley, Gabriel Kreiman, and Michael E Greenberg. Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465(7295):182–7, May 2010.

[106] Martin J Lercher, Araxi O Urrutia, and Laurence D Hurst. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature genetics*, 31(2):180–3, June 2002.

[107] Guo-Ling Zhou, Li Xin, Wei Song, Li-Jun Di, Guang Liu, Xue-Song Wu, De-Pei Liu, and Chih-Chuan Liang. Active chromatin hub of the mouse alpha-globin locus forms in a transcription factory of clustered housekeeping genes. *Molecular and cellular biology*, 26(13):5096–105, July 2006.

[108] Alexey a Gavrilov, Inna S Zukher, Elena S Philonenko, Sergey V Razin, and Olga V Iarovaia. Mapping of the nuclear matrix-bound chromatin hubs by a new M3C experimental procedure. *Nucleic acids research*, 38(22):8051–60, December 2010.

[109] Lusy Handoko, Han Xu, Guoliang Li, Chew Yee Ngan, Elaine Chew, Marie Schnapp, Charlie Wah Heng Lee, Chaopeng Ye, Joanne Lim Hui Ping, Fabianus Mulawadi, Eleanor Wong, Jianpeng Sheng, Yubo Zhang, Thompson Poh, Chee Seng Chan, Galih Kunarso, Atif Shahab, Guillaume Bourque, Valere Cacheux-Rataboul, Wing-Kin Sung, Yijun Ruan, and Chia-Lin Wei. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature genetics*, 43(7):630–8, January 2011.

[110] Felix Mitelman, Bertil Johansson, and Fredrik Mertens. The impact of translocations and gene fusions on cancer causation. *Nature reviews. Cancer*, 7(4):233–45, April 2007.

[111] Johnson Alberts Lewis, Raff, Roberts, Walter. *Molecular Biology of the Cell*. Garland Science, 5 edition, 2007.

[112] Maria Pia Longhese, Davide Mantiero, and Michela Clerici. The cellular response to chromosome breakage. *Molecular microbiology*, 60(5):1099–108, June 2006.

[113] Rüdiger Hehlmann, Andreas Hochhaus, and Michele Baccarani. Chronic myeloid leukaemia. *Lancet*, 370(9584):342–50, July 2007.

[114] C G Geary. The story of chronic myeloid leukaemia. *British journal of haematology*, 110(1):2–11, July 2000.

[115] PC Nowell and DA Hungerford. A minute chromosome in human chronic granulocytic leukemia. *Science*, 132:1497–501, 1960.

[116] J D Rowley. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*, 243(5405):290–3, June 1973.

[117] J D Rowley. Chromosome translocations: dangerous liaisons revisited. *Nature reviews. Cancer*, 1(3):245–50, December 2001.

[118] M W Deininger, John M Goldman, and Junia V Melo. The molecular biology of chronic myeloid leukemia. *Blood*, 96(10):3343–56, November 2000.

[119] K Tanaka, M Takechi, J Hong, C Shigeta, N Oguma, N Kamada, Y Takimoto, A Kuramoto, H Dohy, and T Kyo. 9;22 translocation and bcr rearrangements in chronic myelocytic leukemia patients among atomic bomb survivors. *Journal of radiation research*, 30(4):352–8, December 1989.

[120] A Corso, M Lazzarino, E Morra, S Merante, C Astori, P Bernasconi, M Boni, and C Bernasconi. Chronic myelogenous leukemia and exposure to ionizing radiation–a retrospective study of 443 patients. *Annals of hematology*, 70(2):79–82, February 1995.

[121] Ahmedin Jemal, Rebecca Siegel, Elizabeth Ward, Yongping Hao, Jiaquan Xu, Taylor Murray, and Michael J Thun. Cancer statistics, 2008. *CA: a cancer journal for clinicians*, 58(2):71–96, 2009.

[122] E Matutes, R Morilla, N Farahat, F Carbonell, J Swansbury, M Dyer, and D Catovsky. Definition of acute biphenotypic leukemia. *Haematologica*, 82(1):64–6, 1997.

[123] S Ziemin-van der Poel, N R McCabe, R Espinosa 3rd, H J Gill, Y Patel, A Harden, P Rubinelli, S D Smith, M M LeBeau, J D Rowley, and Et al. Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proceedings of the National Academy of Sciences of the United States of America*, 88(23):10735–10739, 1991.

[124] A Biondi, G Cimino, R Pieters, and C H Pui. Biological and therapeutic aspects of infant leukemia. *Blood*, 96(1):24–33, 2000.

[125] Andrei V Krivtsov and Scott a Armstrong. MLL translocations, histone modifications and leukaemia stem-cell development. *Nature reviews. Cancer*, 7(11):823–33, 2007.

[126] C S Chen, P H Sorensen, P H Domer, G H Reaman, S J Korsmeyer, N A Heerema, G D Hammond, and J H Kersey. Molecular rearrangements on chromosome 11q23 predominate in infant acute lymphoblastic leukemia and are associated with specific biologic variables and poor outcome. *Blood*, 81(9):2386–2393, 1993.

[127] Scott a Armstrong, Jane E Staunton, Lewis B Silverman, Rob Pieters, Monique L den Boer, Mark D Minden, Stephen E Sallan, Eric S Lander, Todd R Golub, and Stanley J Korsmeyer. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nature genetics*, 30(1):41–7, 2002.

[128] H J Sutherland, A Blair, and R W Zapf. Characterization of a hierarchy in human acute myeloid leukemia progenitor cells. *Blood*, 87(11):4754–4761, 1996.

[129] D Bonnet and J E Dick. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine*, 3(7):730–737, 1997.

[130] Jolanta Libura, Maureen Ward, Joanna Solecka, and Christine Richardson. Etoposide-initiated MLL rearrangements detected at high frequency in human primitive hematopoietic stem cells with in vitro and in vivo long-term repopulating potential. *European journal of haematology*, 81(3):185–95, September 2008.

[131] Richard K Burt, Yvonne Loh, William Pearce, Nirat Beohar, Walter G Barr, Robert Craig, Yanting Wen, Jonathan a Rapp, and John Kessler. Clinical applications of blood-derived and marrow-derived stem cells for nonmalignant diseases. *JAMA : the journal of the American Medical Association*, 299(8):925–36, 2008.

[132] A W Hamburger and S E Salmon. Primary bioassay of human tumor stem cells. *Science*, 197(4302):461–463, 1977.

[133] T Reya, S J Morrison, M F Clarke, and I L Weissman. Stem cells, cancer, and cancer stem cells. *Nature*, 414(6859):105–11, November 2001.

[134] B D Yu, J L Hess, S E Horning, G A Brown, and S J Korsmeyer. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature*, 378(6556):505–508, 1995.

[135] J L Hess, B D Yu, B Li, R Hanson, and S J Korsmeyer. Defects in yolk sac hematopoiesis in Mll-null embryos. *Blood*, 90(5):1799–1806, 1997.

[136] Craig D Jude, Leslie Climer, Diyong Xu, Erika Artinger, Jill K Fisher, and Patricia Ernst. Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. *Cell stem cell*, 1(3):324–37, September 2007.

[137] N J Zeleznik-Le, A M Harden, and J D Rowley. 11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. *Proceedings of the National Academy of Sciences of the United States of America*, 91(22):10610–10614, 1994.

[138] M Birke, S Schreiner, M P Garcia-Cuellar, K Mahr, F Titgemeyer, and R K Slany. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Res*, 30(4):958–965, 2002.

[139] T A Milne, Y Dou, M E Martin, H W Brock, R G Roeder, and J L Hess. MLL associates specifically with a subset of transcriptionally active target genes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(41):14765–14770, 2005.

[140] Robert K Slany, Catherine Lavau, and Michael L Cleary. The oncogenic capacity of HRX-ENL requires the transcriptional transactivation activity of ENL and the DNA binding motifs of HRX. *Molecular and cellular biology*, 18(1):122–9, January 1998.

[141] Bernd B Zeisig, Tom Milne, María-Paz García-Cuéllar, Silke Schreiner, Mary-ellen Martin, Uta Fuchs, Arndt Borkhardt, Sumit K Chanda, John Walker, Richard Soden, Jay L Hess, and Robert K Slany. Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. *Molecular and cellular biology*, 24(2):617–28, January 2004.

[142] J L Huret, P Dessen, and A Bernheim. An atlas of chromosomes in hematological malignancies. Example: 11q23 and MLL partners. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 15(6):987–9, June 2001.

[143] C Meyer, B Schneider, S Jakob, S Strehl, A Attarbaschi, S Schnittger, C Schoch, M W J C Jansen, J J M van Dongen, M L den Boer, R Pieters, M-G Ennas, E Angelucci, U Koehl, J Greil, F Griesinger, U Zur Stadt, C Eckert, T Szczepanski, F K Niggli, B W Schäfer, H Kempski, H J M Brady, J Zuna, J Trka, L L Nigro, A Biondi, E Delabesse, E Macintyre, M Stanulla, M Schrappe, O a Haas, T Burmeister, T Dingermann, T Klingebiel, and R Marschalek. The MLL recombinome of acute leukemias. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K*, 20(5):777–84, May 2006.

[144] Y Gu, H Alder, T Nakamura, S A Schichman, R Prasad, O Canaani, H Saito, C M Croce, and E Canaani. Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved in acute leukemia. *Cancer Research*, 54(9):2327–2330, 1994.

[145] P a Sung, J Libura, and C Richardson. Etoposide and illegitimate DNA double-strand break repair in the generation of MLL translocations: new insights and new questions. *DNA repair*, 5(9-10):1109–18, 2006.

[146] S Scharf, J Zech, A Bursen, D Schraets, P L Oliver, S Kliem, E Pfitzner, E Gillert, T Dingermann, and R Marschalek. Transcription linked to recombination: a gene-internal promoter coincides with the recombination hot spot II of the human MLL gene. *Oncogene*, 26(10):1361–1371, 2007.

[147] P L Strissel, R Strick, J D Rowley, and N J Zeleznik-Le. An in vivo topoisomerase II cleavage site and a DNase I hypersensitive site colocalize near exon 9 in the MLL breakpoint cluster region. *Blood*, 92(10):3793–3803, 1998.

[148] A Khobta, C Carlo-Stella, and G Capranico. Specific histone patterns and acetylase/deacetylase activity at the breakpoint-cluster region of the human MLL gene. *Cancer Research*, 64(8):2656–2662, 2004.

[149] G R Buck and E L Zechiedrich. DNA disentangling by type-2 topoisomerases. *Journal of Molecular Biology*, 340(5):933–939, 2004.

[150] LF Liu and JC Wang. Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences*, 84(20):7024, 1987.

[151] Irene Collins, Achim Weber, and David Levens. Transcriptional consequences of topoisomerase inhibition. *Molecular and cellular biology*, 21(24):8437–51, December 2001.

[152] Ricky S Joshi, Benjamin Piña, and Joaquim Roca. Topoisomerase II is required for the production of long Pol II gene transcripts in yeast. *Nucleic acids research*, (17):1–9, June 2012.

[153] Bong-Gun Ju, Victoria V Lunyak, Valentina Perissi, Ivan Garcia-Bassets, David W Rose, Christopher K Glass, and Michael G Rosenfeld. A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. *Science (New York, N.Y.)*, 312(5781):1798–802, 2006.

[154] D A Burden and N Osheroff. Mechanism of action of eukaryotic topoisomerase II and drugs targeted to the enzyme. *Biochimica et Biophysica Acta*, 1400(1-3):139–154, 1998.

[155] A M Arnold and J M Whitehouse. Etoposide: a new anti-cancer agent. *Lancet*, 2(8252):912–915, 1981.

[156] Jolanta Libura, Diana J Slater, Carolyn a Felix, and Christine Richardson. Therapy-related acute myeloid leukemia-like MLL rearrangements are induced by etoposide in primary human CD34+ cells and remain stable after clonal expansion. *Blood*, 105(5):2124–31, 2005.

[157] J A Aten, J Stap, P M Krawczyk, C H van Oven, R A Hoebe, J Essers, and R Kanaar. Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. *Science*, 303(5654):92–95, 2004.

[158] B Jakob, J Splinter, M Durante, and G Taucher-Scholz. Live cell microscopy analysis of radiation-induced DNA double-strand break motion. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3172–7, March 2009.

[159] B E Nelms, R S Maser, J F MacKay, M G Lagally, and J H Petrini. In situ visualization of DNA double-strand break repair in human fibroblasts. *Science*, 280(5363):590–592, 1998.

[160] Evi Soutoglou, Jonas F Dorn, Kundan Sengupta, Maria Jasin, Andre Nussenzweig, Thomas Ried, Gaudenz Danuser, and Tom Misteli. Positional stability of single double-strand breaks in mammalian cells. *Nature cell biology*, 9(6):675–82, 2007.

[161] S Kozubek, E Lukásová, A Marecková, M Skalníková, M Kozubek, E Bártová, V Kroha, E Krahulcová, and J Slotová. The topological organization of chromosomes 9 and 22 in cell nuclei has a determinative role in the induction of t(9,22) translocations and in the pathogenesis of t(9,22) leukemias. *Chromosoma*, 108(7):426–35, December 1999.

[162] Luis A Parada, Philip G McQueen, Peter J Munson, and Tom Misteli. Conservation of relative chromosome positioning in normal and cancer cells. *Current biology : CB*, 12(19):1692–7, October 2002.

[163] Louise Harewood, Frédéric Schütz, Shelagh Boyle, Paul Perry, Mauro Delorenzi, Wendy a Bickmore, and Alexandre Reymond. The effect of translocation-induced nuclear reorganization on gene expression. *Genome research*, 20(5):554–64, May 2010.

[164] H Neves, C Ramos, M G Da Silva, A Parreira, and L Parreira. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. *Blood*, 93(4):1197–1207, February 1999.

[165] Jeffrey J Roix, Philip G McQueen, Peter J Munson, Luis a Parada, and Tom Misteli. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nature genetics*, 34(3):287–91, July 2003.

[166] Ian G Cowell, Zbyslaw Sondka, Kayleigh Smith, Ka Cheong Lee, Catriona M Manville, Malgorzata Sidorczuk-Lesthuruge, Holly Ashlene Rance, Kay Padget, Graham Hunter Jackson, Noritaka Adachi, and Caroline a

Austin. Model for MLL translocations in therapy-related leukemia involving topoisomerase IIβ-mediated DNA strand breaks and gene proximity. *Proceedings of the National Academy of Sciences of the United States of America*, 109(23):8989–94, June 2012.