Annunziato2008

Chapter 1

Introduction

The nucleus of a human cell contains approximately 2 metres of DNA [**Annunziato2008**], requiring a huge degree of compaction to fit within the nucleus. In addition to storing DNA the nucleus itself is a highly functional organelle, responsible for the huge variety of complex cellular processes. To achieve this, the contents of the nucleus are very precisely organised into sub-nuclear compartments which specialise in processes such as transcription and replication. In addition to the organisation of the nuclear proteome, chromatin itself is thought to be organised. 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 found in the nucleus along with 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 by the difference in staining within the nucleus [**Heitz1928**].

Heterochromatin stains darkly because it remains highly condensed during interphase, typically relocating to the nuclear periphery. 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 may 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 [**Finlan2008**, **Reddy2008**] though this effect appears to be locus specific and is not always observed [**Kumaran2008**]. Euchromatin is the site of most geneic transcription, as shown by the presence of RNA Polymerase and nascent transcriptions found in early fractionation studies [**Chesterton1974**]. 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

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 [**Davey2002**]. 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.

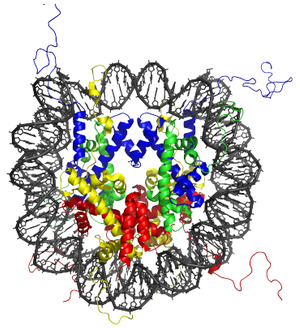

Figure 1.1 – Structure of the nucleosome. X Ray Structure of a Nucleosome core particle at a resolution of 1.9Å. DNA can be seen wrapped around the core histones. PDB structure 1KX5 [**Davey2002**].

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 [**Wolffe1999a**] 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' [**Strahl2000a**], 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), inactive promoters for H3K27me3 and H3K9me3 (Figure 1.2) (for a review, see [**Zhou2011**]).

As chromatin immunoprecipitation (ChIP) has become a common laboratory technique and been combined with microarray techniques (ChIP on chip) and next generation sequencing (ChIP-Seq), our understanding of how these marks 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 [**Heintzman2007**, **Ernst2010**, **Hon2009**]. 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[**Ernst2011**]. 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.

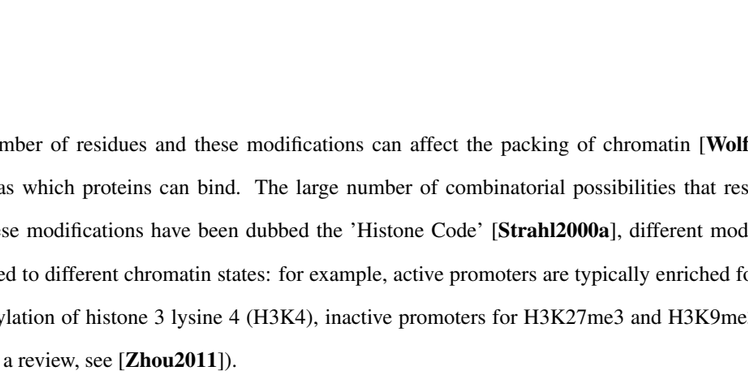

Figure 1.2 – Histone modification associations in chromatin. Diagram showing the range of histone modifications available within mammalian chromatin and how they can specify the behaviour of elements within the DNA. Taken from[**Zhou2011**].

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 extremely tightly conserved between species. Histone variants are found as single genes spread through the genome, and are subject to far greater diversity [**Talbert2010**].

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 [**Dunleavy2009**, **Foltz2009**]. 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 [**Talbert2010**]. This histone variant is found within transcribed genes, promoters and regulatory elements, and is thought to be laid down during transcriptional elongation [**Schwartz2005**]. Nucleosomes containing H3.3 appear to be less stable than canonical nucleosomes, with a high turnover of H3.3 [**Schwartz2005**]. It is possible that this increased turnover of the nucleosomal components helps to keep the chromatin open and accessible to the transcriptional machinery [**Talbert2010**].

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 [**Zlatanova2008**]. H2A.Z promotes the recruitment of RNA Polymerase II through mediating protein interactions [**Adam2001**].

1.2 Two-dimensional organisation

Chromosome banding, *cis* organisation of Hox genes, gene clusters, gene rich areas and gene deserts.

1.3 Techniques to investigate nuclear organisation

1.3.1 Microscopy

1.3.1.1 Chromosome painting

1.3.1.2 DNA-FISH

1.3.1.3 RNA-FISH

1.3.1.4 Electron microscopy

1.3.2 Chromosome conformation capture

Chromosome conformation capture (3C)

1.3.2.1 Circular 4C

1.3.2.2 e4C

1.3.2.3 5C

1.3.2.4 ChIA-PET

1.3.2.5 Hi-C

1.3.2.6 Tethered Hi-C

1.4 Chromosome Territories

The interphase nucleus is a highly structured organelle. As chromosomes decondense after metaphase they retain some degree of structure, forming *chromosome territories* (CTs) [**Cremer2001**]. Circumstantial evidence for interphase organisation of chromosomes has existed for a long time, first suggested by Carl Rabl in 1885 [**Rabl1885**]. Oberservations by Stack *et al.* using microscopy with giemsa-band staining suggested that chromosomes retained some degree of organisation during interphase [**Stack1977**], 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 [**Cremer1982**]. The

subsequent development of 'chromosome paints', a method to visualise entire- or part- chromosomes with fluorescence *in-situ* hybridisation (FISH), confirmed these findings [**Schardin1985a**, **Manuelidis1985**].

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 [**Sun2000**, **Cremer2001b**], gene-density [**Cremer2001b**, **Croft1999**] and replication timing [**Visser1998**]: 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 [**Tanabe2002**] and shown tissue specificity in chromosome pairing [**Parada2004b**].

It's known that for some but not all genes, positioning at the nuclear periphery correlates with reduced gene expression [**Kosak2002**, **Dietzel2004**, **Zink2004a**]. To investigate whether transcriptional activity is the cause or effect of this positioning effect, two groups published studies which artificially tethered genomic regions to the inner nuclear membrane using Lac operators *(lacO)* [**Finlan2008**, **Reddy2008**]. 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 [**Finlan2008**]. Whilst these studies suggest that it is nuclear positioning that leads to transcriptional effects, a study of CT positioning by Croft *et al.* showed that inhibition of transcription causes a reversible change in CT position [**Croft1999**]. Large scale rearrangements of CTs have also been observed during cell differentiation [**Stadler2004**]. Interestingly, derivative chromosomes that result from balanced translocations affect the organisation of CTs within the nucleus [**Harewood2010**], raising the possibility that the global changes in gene expression observed after oncogenic translocations could be in part due to changes in genome organisation [**Harewood2010**].

1.5 Nuclear compartmentalisation

Cajal bodies, speckles, nucleosomes

1.6 Transcription factories

One nuclear subcompartment which has come to light within the past twenty 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 [**Jackson1993a**, **Osborne2004**, **Eskiw2008a**], challenging the classical model of transcription found in many text books.

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 [**Jackson1993a**]. Further studies showed that these foci contain RNA Polymerase II along with many other components required for transcription [**Iborra1996**, **Grande1997**]. 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 associated with the surface of large nitrogen-rich protein structures with a diameter of ~87 nm, comparable in size to that of a transcription factory [**Eskiw2008a**].

The stability of nascent transcripts was shown as far back as 1981 by Jackson and colleagues [**Jackson1981**] and the visualisation of transcription factories has demanded a new model for the action of RNA Polymerase II [**Cook1999**]. The revised model proposes that instead of RNA Polymerase II freely diffusing to active genes and tracking along the gene body, it is the genes that are recruited to transcription factories and the genes that track 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

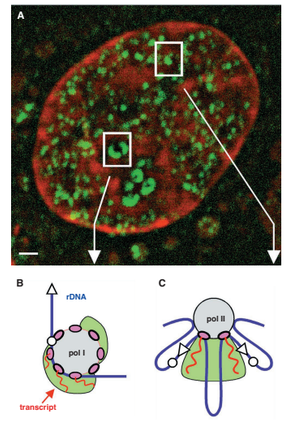

Figure 1.3 – Nascent RNA and transcription factories. (A) Transcription foci in HeLa cells, visualised with labelled Br-UTP in 100 nm cryosections. Nascent RNA (green) is concentrated in punctate foci. (B) Model for a nucleolar factory, showing a transcript with multiple polymerases generating a crescent shaped foci. (C) Model for a nucleoplasmic factory. Multiple transcribed regions each with a single polymerase generate a smaller cloud of nascent transcripts. Taken from Cook *et al.*, Science (2009) [**Cook1999**].

is closer to 1 mM [**Carter2008**]. 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 [**Iborra1996**, **Cook1999**], creating topological loops within the template DNA instead 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 [**Jackson1998**]. They showed that each HeLa cell nucleus contain ~ 2400 transcription factories, each with approximately 30 active RNA-Polymerase II complexes [**Jackson1998**]. 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 [**Osborne2004**]. 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 [**Osborne2004**, **Jackson1998**].

1.6.1 Specialised transcription factories

[**Schoenfelder2010**]

1.7 Looping

1.7.1 Chromatin hubs

Hbb, Igh and other classical looping examples.

1.7.2 Long range interactions

Sonic the hedgehog, other long range *cis* interactions? Any evidence of *trans* stuff? Cam's Myc and Igh FISH paper.

1.8 Domains

Hi-C domains. Interactions within domains. Talk about how transcription depends on context. Spreading of epigenetic marks.

1.9 What drives nuclear organisation?

1.9.1 Actin and myosin

1.9.2 Transcription

Active RNA Polymerases: Mobile or Immobile Molecular Machines? [**Papantonis2010**]

1.9.3 CTCF and cohesin

Jackson, McCreedy and Cook showed in 1981 that nascent RNA transcripts labelled with [^{3}H] uridine remained within the nucleus when loops of DNA were removed using a nuclease, suggesting that transcription occurs close to points of chromatin attachment [**Jackson1981**].

Igh forming rosettes for v region selection - http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3070187/

1.9.4 Transcriptional hubs

Talk about latest theories about housekeeping genes organising everything else. How this ties in with chromatin domains.

1.9.5 The self-organisation model

Cook 2002, Iborra and Cook 2002.

1.10 The annotated genome

Talking about the use of a systems biology approach to understand the linear genomic information of the genome.

1.11 Chromosomal translocations

1.11.1 *BCR* and *ABL1*

1.11.2 Mixed lineage leukaemia

1.11.3 Formation of chromosomal translocations

Basics of HR, SSA and NHEJ. Contact first / breakage first models.

1.11.3.1 Double strand breaks and religation

1.11.3.2 Contact first model

1.11.3.3 Chromosome Territories

Branco and Pombo, Plos Biol 2006

[**Parada2004b**] - tissue specific CT pairing in mouse, correlates with tissue translocation incidence

1.11.3.4 Transcription Factories

Cam's Myc and Igh RNA-FISH paper.

1.12 Hypothesis and aim of thesis

Talk about the why of translocations - specificity and frequency of certain translocations means that partner selection can't be completely random. Hypothesis and aims: that specific gene pairs are predisposed to chromosomal translocations because they are co-localising at transcription factories. Aims of the project are to investigate the genome-wide association networks of specific oncogenes and see if gene pair co-localisation frequency matches up with gene-pair translocation frequency.