

Chapter 1

Introduction

The nucleus of a human cell contains approximately 2 metres of DNA [2], 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 [17].

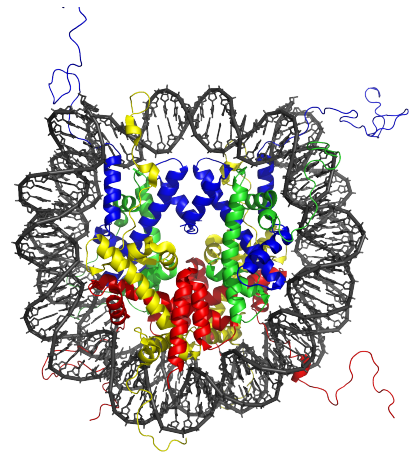
Heterochromatin stains darkly because it remains highly condensed during interphase, typically relocating to the nuclear periphery. can be split into *constitutive heterochromatin* and *facultative heterochromatin*; the former describing heterochromatin found in all cell types typically containing repetitive 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 [13, 26] though this effect appears to be locus specific and is not always observed [23]. Euchromatin is the site of most geneic transcription, as shown by the presence of RNA Polymerase and nascent transcriptions found in early fractionation studies [4]. 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 octamer 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 [8]. 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 backbone.



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 [32] as well as which proteins can bind. The large number of combinatorial possibilities that result from these modifications have been dubbed the 'Histone Code' [28], 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 [33]).

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.

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 [8].

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 [16, 10, 18]. 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 [11]. 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.

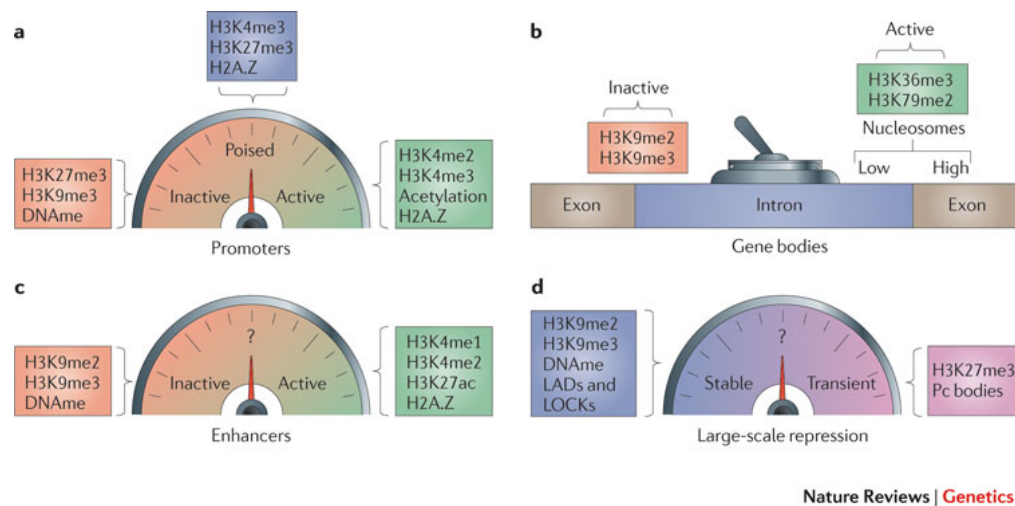


Figure 1.2: Histone modification associations in chromatin

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 [30].

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 [9, 14]. 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 [30]. This histone variant is found within transcribed genes, promoters and regulatory elements, and is thought to be laid down during transcriptional elongation [27]. Nucleosomes containing H3.3 appear to be less stable than canonical nucleosomes, with a high turnover of H3.3 [27]. It is possible that this

increased turnover of the nucleosomal components helps to keep the chromatin open and accessible to the transcriptional machinery [30].

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 [34]. H2A.Z promotes the recruitment of RNA Polymerase II through mediating protein interactions [1].

1.2 Two-dimensional organisation

cis organisation of Hox genes, gene clusters, gene rich areas and gene deserts

1.3 Chromosome Territories

Chromosome painting, what they are, not much more. (talk about organisation of CTs later). CTs being close to the nuclear periphery, CT intermingling, being buried in CTs or on the surface.

Interphase chromosomes are organised into discrete regions, known as chromosome territories (CTs) [6]. Chromosome territories are not randomly positioned, and territory position has been shown to be conserved through evolution [31]. Fluorescence *in-situ* hybridisation (FISH) studies have demonstrated large scale rearrangements of CTs with changes in cell morphology, such as in chicken hematopoietic cell differentiation. Studies of CT localisation within multiple mouse cell types have shown that CT organisation can be tissue specific [25]. Certain chromosomes have preferred neighbours in some tissue types, for example chromosomes 12 and 15 are frequently close in mouse lymphocytes but not mouse hepatocytes [25]. This is relevant as translocations between these chromosomes are common in lymphomas but not in hepatomas. The positions adopted by chromosome territories within the nucleus appears to be related to the sequence encoded by each chromosome – those near the centre of the nucleus tend to be gene-rich, often replicate early and are frequently relatively small, whilst the opposite is true for chromosomes located near the nuclear periphery [29, 7].

1.4 Nuclear compartmentalisation

Cajal bodies, speckles, nucleosomes

1.5 Transcription factories

Transcription factories in detail as distinct sub-compartments. Transcription takes place in transcription factories. Stochastic bursting transcription.

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 [22, 24, 12], 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 prescence of the RNA Polymerase II inhibitor α -amanitin [22]. Further studies showed that these foci contain RNA Polymerase II along with many other components required for transcription [19, 15]. 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 visualize 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 [12].

The stability of nascent transcripts was shown as far back as 1981 by Jackson and colleagues [21] and the visualisation of transcription factories has demanded a new model for the action of RNA Polymerase II [5]. 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 μ M concentration of active RNA Polymerase II, whereas the local concentration within transcription factories is closer to 1 mM [3]. 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

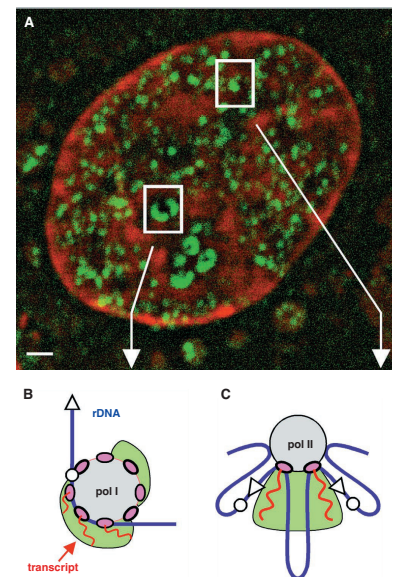


Figure 1.3: Nascent RNA and transcription factories.

(A) Transcription foci in HeLa cells, visualised with labelled Br-UTP in 100 nm cryosections. Nascent RNA (green) is concentrated in punctate

factories would extrude their RNA transcripts into the nucleoplasm [19, 5], 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 [20]. They showed that each HeLa cell nucleus contain ~ 2400 transcription factories, each with approximately 30 active RNA-Polymerase II complexes [20]. 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 [24].

Osborne et al. used DNA-FISH (fluorescence in-situ hybridisation), RNA-FISH, immunofluorescence and 3C (chromosome conformation capture) to show that genes situated within transcription factories are actively transcribed, whereas those outside are not [2]. Transcription was shown to be discontinuous, 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, confirming earlier predictions that genes must share transcription factories [2, 18]. Osborne et al. suggested that the nuclear architecture may be in part organised by transcription.

The transcription factory model provides answers to several theoretical problems regarding transcription: a static polymerase with nearby topoisomerase activity would allow nascent RNA to extrude freely from the transcript and avoid tangling with the DNA. Localised centres of transcription may associate with other specialised complexes, such as specific transcription factors and splicing components. Indeed, a great deal of research has gone into testing these predictions; many of the proteins involved in capping, splicing and 3' processing of nascent mRNA, processes that take place co-transcriptionally, have been shown to interact with the carboxy-terminal domain (CTD) of RNA Polymerase II [20], prompting one author to suggest that transcription factories should be renamed 'mRNA factories' [21]. Transcription factory specialisation may have further implications in the epigenetic control of gene expression: if transcription of a gene only occurs within a limited number of factories in the nucleus, the epigenetic control of chromatin mobility may affect the ability of that gene to relocate to its transcription factory.

1.5.1 Specialised transcription factories

1.6 Looping

1.6.1 Chromatin hubs

Hbb, Igh and other classical looping examples.

1.6.2 Long range interactions

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

1.7 Domains

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

1.8 What drives nuclear organisation?

1.8.1 Actin and myosin

1.8.2 Transcription

1.8.3 CTCF and cohesin

Jackson, McCreedy and Cook showed in 1981 that nascent RNA transcripts labelled with [³H] uridine remained within the nucleus when loops of DNA were removed using a nuclease, suggesting that transcription occurs close to points of chromatin attachment [21].

1.8.4 Transcriptional hubs

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

1.8.5 The self-organisation model

Cook 2002, Iborra and Cook 2002.

1.9 The annotated genome

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

1.10 Techniques to investigate nuclear organisation

1.10.1 Microscopy

1.10.1.1 Chromosome painting

1.10.1.2 DNA-FISH

1.10.1.3 RNA-FISH

1.10.1.4 Electron microscopy

1.10.2 Chromosome conformation capture

Chromosome conformation capture (3C)

1.10.2.1 Circular 4C

1.10.2.2 e4C

1.10.2.3 5C

1.10.2.4 ChIA-PET

1.10.2.5 Hi-C

1.10.2.6 Tethered Hi-C

1.11 Chromosomal translocations

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

1.11.1 Formation of chromosomal translocations

1.11.1.1 Double strand breaks

1.11.1.2 Aberrant religation

1.11.1.3 Contact first model

1.11.2 *BCR* and *ABL1*

1.11.3 Mixed lineage leukaemia

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.

Bibliography

- [1] 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.
- [2] A. Annunziato. DNA Packaging: Nucleosomes and Chromatin. *Nature Education*, 1(1), 2008.
- [3] David R F Carter, Christopher Eskiw, and Peter R Cook. Transcription factories. *Biochemical Society transactions*, 36(Pt 4):585–9, August 2008.
- [4] 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.
- [5] P. R. Cook. The Organization of Replication and Transcription. *Science*, 284(5421):1790–1795, June 1999.
- [6] T Cremer and C Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2(4):292–301, 2001.
- [7] 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.
- [8] 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):1097–1113, June 2002.
- [9] 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.

- [10] 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.
- [11] Jason Ernst, Pouya Kheradpour, Tarjei S Mikkelsen, Noam Shores, 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] E. Heitz. Das heterochromatin der moose. *I Jahrb Wiss Botanik*, 69:762–818, 1928.
- [18] 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.
- [19] 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.
- [20] 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.

- [21] D A Jackson, S J McCready, and P R Cook. RNA is synthesized at the nuclear cage. *Nature*, 292(5823):552–555, 1981.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] Luis a Parada, Philip G McQueen, and Tom Misteli. Tissue-specific spatial organization of genomes. *Genome biology*, 5(7):R44, January 2004.
- [26] 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.
- [27] 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.
- [28] B D Strahl and C D Allis. The language of covalent histone modifications. *Nature*, 403(6765):41–5, January 2000.
- [29] H B Sun, J Shen, and H Yokota. Size-dependent positioning of human chromosomes in interphase nuclei. *Biophysical Journal*, 79(1):184–190, 2000.
- [30] 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.
- [31] 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.
- [32] a P Wolffe and J J Hayes. Chromatin disruption and modification. *Nucleic acids research*, 27(3):711–20, February 1999.
- [33] 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.

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