Chapter from the book *DNA Replication - Current Advances*

Downloaded from: [http://www.intechopen.com/books/dna-replication-current-advances](http://www.intechopen.com/books/dna-replication-current-advances)

Interested in publishing with IntechOpen?
Contact us at book.department@intechopen.com
Replication Timing: Evolution, Nuclear Organization and Relevance for Human Disease

Megan Lynne Wright and Frank Grützner
The Robinson Institute, School of Molecular & Biomedical Sciences, University of Adelaide, Adelaide Australia

1. Introduction
DNA replication in eukaryotes is multifaceted, dynamic and highly organised. In contrast to bacterial cells, which replicate from single origins of replication, complex eukaryote genomes replicate from thousands of origins of replication. Although we know that the timing of replication depends on the chromatin environment, the function and evolution of mechanisms controlling replication timing are unclear. Many studies in species ranging from yeast to humans have demonstrated how replication timing depends on proximity to certain sequences such as telomeres and centromeres (Ferguson and Fangman, 1992; Friedman et al., 1996; Heun et al., 2001), chromatin status (euchromatin and heterochromatin) and is linked to gene function and expression (housekeeping genes versus tissue specific genes and monoallelically expressed genes) (Hiratani and Gilbert, 2009; Hiratani et al., 2009). Replication timing has been linked to fundamental epigenetic regulatory mechanisms including genomic imprinting (Kitsberg et al., 1993; Knoll et al., 1994), X chromosome inactivation (Gilbert, 2002; Takagi et al., 1982; Wutz and Jaenisch, 2000), interchromosomal interactions (Ryba et al. 2010) and is increasingly recognised to be important in human disease (DePamphilis, 2006).

This chapter integrates established knowledge with recent scientific breakthroughs, using genome-wide approaches linking different aspects of epigenetic control with replication timing, to provide a state-of-the-art overview and perspective for future work in this area of research. Despite detailed knowledge on replication timing in a select number of model organisms (e.g. yeast, drosophila, mouse) we are only beginning to understand how replication timing evolved in relation to other epigenetic mechanisms (e.g. genomic imprinting, X inactivation, and long-range chromatin interaction). The evolution of these epigenetic mechanisms will be presented together with novel ideas about how cytological and genome-wide approaches and methodologies can be combined to provide a comprehensive picture of spatial and temporal organization, the evolution of replication timing in eukaryotic genomes, and their relevance in human disease.

2. Background
2.1 Replication initiation
The complete and accurate replication of DNA during the S-phase is of fundamental importance for all organisms. The mechanism of replication is highly conserved across
evolution, whereby a cell must gather the proteins to initiate replication at specific origins of replication (ORs), unwind the DNA, move the replication fork bi-directionally away from the OR in such a manner as to allow the replication of the new daughter strand of DNA using the old parental DNA strand, and then cease replication. However whilst the replication process is highly conserved, different eukaryotes use different proteins and forms of control over replication (Gilbert, 2010).

Whilst general similarities exist in the type of machinery required to copy and create a new DNA strand across organisms, some areas of genome replication remain elusive. One such area in eukaryotes is replication initiation and timeline. Linear eukaryotic chromosomes replicate from many ORs which are spread out along their structure and are recognized by the origin recognition complex (ORC) (reviewed in Masai et al., 2010). These OR sites are where replication forks form and move bi-directionally away from the OR, replicating the DNA sequence as they move, then terminating when they meet another fork approaching from the opposite direction. The ORCs recognize almost all ORs, and will assemble at these regions in a highly conserved manner across eukaryotes. However, whilst ORCs bind specific sequence motifs in some eukaryotes, such as in budding yeast (Bell and Stillman, 1992), in other eukaryotes specificity is not well defined through sequence. Fission yeast and Drosophila have ORCs that recognize AT-rich sequences (Austin et al., 1999; Chuang and Kelly, 1999), rather than specific motifs. Moreover, human ORCs, which are chosen as initiators of replication, have also been shown to require AT-rich sequences as well as various other features, including matrix attachment region sequences, dinucleotide repeats and asymmetrical purine-pyrimidine sequences (Altman and Fanning, 2004; Debatisse et al., 2004; Paixao et al., 2004; Schaarschmidt et al., 2004; Wang et al., 2004). Other factors that may affect the initiation of replication at certain ORs also include DNA topology, transcription factors, and elements of the pre-replicative complex (pre-RC) (reviewed in Masai et al., 2010).

During late mitosis and G1, the chromatin-bound ORCs are loaded with minichromosome maintenance (MCM) complex, and thus become pre-RCs, with the ability to gather the required components to start replication. The pre-RCs assemble at most of the OR regions, however only a few of these complexes start replication in their region. The cell’s choice to start replication at some ORs as opposed to others is unclear; whilst it is thought that the assembly of the pre-RCs at most ORs is used as backup in case the cell runs into trouble during replication, the choice as to whether a Pre-RC becomes an active replication initiator is not well understood (Doksani et al., 2009; Ibarra et al., 2008; Koren et al., 2010; Woodward et al., 2006).

There are, however, some known factors that may contribute to a pre-RC site becoming an active OR (reviewed in Masai et al., 2010). Firstly, the selection of replication initiation sites may be controlled by both the existence of a pre-RC and its assembly in combination with events that actually cause initiation. For example, the firing of an OR appears to affect the firing of adjacent ORs, as shown in the example of budding yeast, where active ORs suppress the initiation of replication at adjacent ORs (Brewer and Fangman, 1993). In this example, the suppression of adjacent potential ORs may be caused by the disruption of pre-RC complexes at these sites by the replication process initiated at the active OR (reviewed in Masai et al., 2010). Also, read-through transcription may affect the firing of downstream ORs (Haase et al., 1994; Saha et al., 2004). Furthermore chromatin structure, which refers to the chemical characteristics of the chromatin strand, may influence the initiation of replication by affecting the pre-RC assembly. There is evidence to show that histone
acetylases and deacetylases play roles in the assembly of pre-RCs by interacting with, or disturbing the loading of, pre-RC elements such as the MCM complex (Burke et al., 2001; Iizuka et al., 2006; Pappas et al., 2004; Pasero et al., 2002).

Finally, distal elements, such as locus control regions (LCRs) are known to affect initiation (Hayashida et al., 2006; Kalejta et al., 1998), with the initiation of replication at regions such as the human $\beta$-globin locus being controlled by a $5'$ LCR (Aladjem et al., 1995).

### 2.2 Temporal programmes of ORs in eukaryotic chromosomes

Replication of eukaryotic genomes follows a defined temporal program, whereby the firing of ORs occurs in a predetermined but tissue specific manner. Hence this process is dynamic in terms of the selection of OR activation, as the cellular environment also plays a role in the temporal regulation of replication across the genome. Experiments have shown that a reduction in cellular thymidine caused a reduction in replication fork speed. This caused more intermediate ORs to be activated in order to compensate for the reduction in replication speed (Anglana et al., 2003; Taylor, 1977), and showed that cellular environments indeed affect the dynamics of OR firing. This shows that a cell is able to change its predetermined temporal replication program if it undergoes replication stress, with the most relevant aspect of OR activation being the genomic context and how it impacts the replication program.

Factors that are involved in OR firing include chromatin loops, dormant and active pre-RC complexes and fork replication rate, and finally nuclear organisation. Firstly, there is some evidence to suggest that chromatin loops affect replication firing. Studies in *Xenopus* egg extracts transferred with erythrocyte nuclei showed that cells that entered into M-phase instantly after somatic transfer took longer to replicate than cells which were held in mitosis and allowed to undergo a single mitosis event. This was due to the influence of the single round of mitosis on the chromatin structure; the round of mitosis supported the formation of smaller chromatin loops which correlated with higher ORC protein recruitment and more efficient genome replication (Lemaitre et al., 2005). Another study showed that the ORs closer to regions of chromatin loop anchorage in G1 initiated replication in the following S-phase earlier than ORs located further away from anchorage regions, indicating that loop-formation was part of the control mechanism for OR firing (Courbet et al., 2008).

Fork replication rate also appears to have a role in the temporal organization of OR firing. Genomic integrity may be aided by the presence of dormant origins of replication, as MCMs are often present in much greater amounts than those needed at pre-RCs, and the reduced presence or loss of pre-RCs result in genomic instability, S-phase arrest, and cell death (Edwards et al., 2002; Hyrien et al., 2003; Lengronne and Schwob, 2002; Shreeram et al., 2002; Tanaka and Diffley, 2002). Dormant ORs have been shown to activate when forks are stalled, with one model hypothesizing that OR activation occurs stochastically, whereby the presence of a stalled fork increases the chances of adjacent dormant ORs being activated (Blow and Ge, 2009; Ibarra et al., 2008). Other models propose that the presence of a stalled fork changes the topology of the DNA strand and the chromatin structure within the region, thus causing nearby and usually dormant ORs to activate (Ibarra et al., 2008).

Finally, nuclear organisation has a role to play in a cell’s replication program. Distinct chromosome territories exist as separate nuclear architecture compartments in interphase cells. Within these territories, a higher order of chromatin structure exists, where domains containing specific chromosomal arms and bands have been found to be located in the
nucleus in similar regions of certain cell types (Dietzel et al., 1998). It has also been proposed that these chromatin-rich chromosome territories (CTs) are separated by chromatin-poor areas called ‘interchromatin compartments’, which contain transcriptional and splicing machinery, as well as DNA replication and damage-repair machinery (reviewed in Aten and Kanaar, 2006; Cremer and Cremer, 2001; Misteli, 2001). However recent work showed extensive intermingling of CTs contradicting the existence of the interchromatin compartment (reviewed in Aten and Kanaar, 2006; Branco and Pombo, 2006; reviewed in Cremer and Cremer, 2010). Within separate chromosome territories there are many replication foci, whereby early and late replicating DNA can be found in spatially separate and distinct regions (Zink et al., 1999). Overall late replicating DNA (including the late replicating inactive X chromosome) is often located at the nuclear periphery or around the nucleolus organizing region (Sadoni et al., 1999).

3. Asynchronous replication
Asynchronous replication is another variation in the eukaryotic temporal replication repertoire. Asynchronous replication occurs when the ORs present in the same regions on two homologous chromosomes, initiate replication at different times. This results in one of the alleles replicating earlier than the allele on the other homologue. Notably, the alleles of asynchronously replicating genes are also observed to locate to separate discrete foci in a nucleus. This form of replication is a feature of monoallelically expressed genes, including genes that undergo allelic exclusion, imprinted genes, and genes from the X-chromosome in female somatic cells.

3.1 Approaches to measuring asynchronous replication and its effects on genome biology and disease
3.1.1 Chromosome banding
Chromosome banding techniques gave the first insights into the epigenetics behind replication, and more specifically, asynchronous replication. It is now well established that replication timing is not uniform across eukaryotic genomes, with select chromosomal regions showing early or late replication in the S-phase. This phenomenon has been observed in distinct banding regions along condensed metaphase chromosomes. The discovery of early and late replication banding on metaphase chromosomes using the Bromodeoxyuridine (BrdU) incorporation technique, can be attributed to Latt (1973). Latt discovered that the differential incorporation of BrdU, a thymidine replacement, during the S-phase between early and late replicating regions of DNA, could be measured using 33258 Hoechst fluorescence. An efficiency reduction of the Hoechst dye fluorescence occurs when it is bound to the incorporated poly(dA-BrdU) compared to the poly(dA-dT). Incorporation of BrdU into either the late or early replicating DNA can be adjusted by culturing cells in BrdU for different time periods; specifically early replication stage BrdU incorporation was achieved by first culturing in BrdU with the addition of a terminal pulse of [3H]-dT, whilst late replication BrdU incorporation was achieved by culturing in medium containing thymidine to which BrdU was only added 6 hours before harvest. This allowed identification of 5-10 megabasepair regions on chromosomes replicating either early or late in the S-phase.

Latt’s early research defined a fundamental relationship between chromosome organisation and replication timing; eukaryotic chromosomes do not undergo equivalent amounts of
replication both within a chromosome and across a karyotype, whereby a distinct non-equivalence of replication is represented by the presence of discrete bands for early and late replicating regions on a chromosome. Furthermore, the late-replicating inactive X chromosome in human females, which is noted to have a slightly more condensed karyotype, showed distinctly opposing fluorescence to the less-condensed active-X chromosome.

Higher-resolution replication banding has since been established in humans and numerous vertebrate species (Biederman and Lin, 1979; Costantini and Bernardi, 2008; Drets et al., 1978). Currently, there are three tiers of replication resolution: 1) low-resolution banding (e.g. De Latt’s BrdU bands, and Giemsa and Quinacrine bands); 2) higher resolution banding (GC content in grouped isochore regions); and 3) individual isochores (Costantini and Bernardi, 2008). Isochores are regions of DNA, above 300 kb (on average around 0.9 Mb in size in the human genome), that have a similar GC content, and also have similar gene content (Costantini and Bernardi, 2008; Costantini et al., 2006; Costantini et al., 2007). Specifically, there are five groups of isochores, whereby lower GC content is classed with the isochore groups L1 and L2 (less than 40% GC-content, and few genes), intermediate groups are H1 and H2 (with around 47% and 52% GC-content, and intermediate amounts of genes), and finally the highest group is H3 (with above 52% GC-content, and high amounts of genes) (Bernardi, 1995). A replicon is a genomic region around 50-400 kb in size, that replicates from a single origin of replication. It has been shown that replicons that exist within a certain isochore region, all undergo similar replication timing, with clusters of early replicating replicons being found next to each other, and clusters of late-replicating replicons being grouped as well (Watanabe et al., 2002). Through the comparisons of the three tiers of resolution, it was found that groups of early and late replicating isochores corresponded to, and approached the same size of, high-resolution replication banding regions (4-7 Mb).

The results of the highest-replication isochore banding when compared to the other banding techniques has indicated that in mammalian chromosomes there are three nested structures important to replication (Figure 1). The first structure is that of the replicon (50-450 kb), whereby individual replicons undergo dynamic firing of their ORs. These replicons however usually exist in clusters of 10 or more, and every replicon in the cluster will usually undergo replication at the same time during the S-phase. The second is that of the isochore (> 300 kb) which is a region that exists as a combination of replicons all with similar early or late replication status and GC content, which can undergo early or late replication in the cell cycle. The third structure is that of the cytogenetic bands, which indicates large regions on a chromosome undergoing early or late replication, and corresponds well to groups of all-early or all-late replicating isochores (Costantini and Bernardi, 2008). This shows that the arrangement of mitotic chromosome structure is closely related to replication timing, from the chromosome banding level, all the way through to the level of organisation of the individual replicons. This pattern is maintained in interphase, where chromosome territories in the S-phase have clusters of early and late replicating foci, which correspond to the R- and G/C bands observed in mitotic chromosomes respectively (Sadoni et al., 1999).

Replication banding techniques have allowed early and late timing replication zones to be delineated along metaphase chromosomes, where areas of similarly replicating replicons are grouped making larger replicon clusters (Watanabe et al., 2002). However, the large genomic regions that bridge the transition of an early-replicating replicon cluster to a late-replicating replicon cluster appear to lack any ORs, and rely on the continuous movement of
forks from adjacent replicon-clusters/isocho re regions for replication to occur in their region (reviewed in Farkash-Amar et al., 2008; Hiratani et al., 2008; Watanabe and Maekawa, 2010). This means that the fork from the earlier firing OR will have to move across the replication transition region, until it meets another fork from the late-replicating region. This will often pause replication in these early to late transition zones, which can cause genomic instability in the form of DNA breaks and rearrangements (Raghuraman et al., 2001; Rothstein et al., 2000). Furthermore, common genomic fragile sites frequently reside in early to late replication transition regions, and also lack backup ORs (Debatisse et al., 2006; Ge et al., 2007; Ibarra et al., 2008).

In addition to the increased genomic instability there is also an increase in the number of non-B-form DNA structures in replication transition regions (reviewed in Watanabe and Maekawa, 2010). Replication switch points (from early to late) are often associated with purine/pyrimidine rich areas, as these DNA regions can form structures called triplexes (H-DNA) that are known stop replication forks (Baran et al., 1991; Brinton et al., 1991; Ohno et al., 2000). The non-B-form structures however also have mutagenic properties, causing somatic recombination events (Kalish and Glazer, 2005; Knauert et al., 2006). It has thus been
proposed that these replication transition regions, which correspond to the regions between R/G bands, are subject to more genomic instability due to the increased presence of non-B-DNA structures in these genomic areas (Watanabe and Maekawa, 2010).

Replication timing is affected in regions of the human genome involved in disease. Generally it has been proposed that regions of the human genome that reside in areas where replication timing switches (early to late) would be unstable and more prone to DNA damage (reviewed in Watanabe and Maekawa, 2010). Notably, these regions of replication timing transition are also associated with many human diseases, including cancer (Watanabe et al., 2009; Watanabe et al., 2002; Watanabe et al., 2004). Regions or genes associated with other diseases, such as familial Alzheimer’s, familial amyotrophic lateral sclerosis and phenylketonuria, are also found in these replication timing transition regions. Furthermore, there are over 70 human diseases associated with non-B DNA structures, including neurological and psychiatric diseases, and many genomic disorders, indicating that the increase of these structures in replication timing transition regions may be a first step in the mutational process associated with these diseases (reviewed in Watanabe and Maekawa, 2010).

3.1.2 Measuring asynchronous replication with the dot assay technique

Molecular cytogenetic techniques like Fluorescence in situ Hybridization (FISH) and an explosion of available genomic clones and whole chromosome probes has let to huge refinement of physical maps on metaphase and interphase chromosomes. This also enabled replication timing to be investigated on the single gene level. In these experiments, DNA probes designed to hybridise to a specific gene allowed the replication status to be observed in three states in a nucleus; two signals (single-single (SS) dot) represents an unreplicated status, whilst a three signal status (single-double (SD) dot) represents a locus undergoing replication. The fourth state (single-single (DD) dot) represents that both the alleles have replicated, and the locus has finished replicating.

Fig. 2. Cytogenetic FISH dot assay

Mammalian interphase nuclei stained with DAPI (blue). Yellow arrows point to allele copies (green FISH signals) observed in each nucleus. The SS panel has two clearly defined green signals representing the two allele copies present in the nucleus, meaning the locus has not replicated. The SD panel shows three green signals, indicating that one allele has undergone replication, whilst the other allele is lagging behind and not yet replicated. The DD panel shows 4 green signals, indicating that both the alleles have replicated, and the locus has finished replicating.
replication, where one allele has replicated and the other is lagging, and finally a four signal status (double-double (DD) dot) represents a locus that is fully replicated (Selig et al., 1992). Asynchrony in this case is measured by the frequency of three-signal (SD dot) status observed in a cell line. However, the classification of asynchronous replication varies in the literature, with an asynchronously replicating state being assigned for loci with anywhere between 30-50% SD signal, and a non-asynchronously replicating locus generally having below 30% SD signal (Baumer et al., 2004; Wilson et al., 2007).

4. Replication timing in heteromorphic sex chromosomes

Replication banding and FISH dot assay techniques have not only shed light on how chromosome structure can affect replication, they have also allowed new insights into how replication timing of single genes has evolved. Changes in replication banding specific to one homolog in a karyotype have been used to identify early stage cytologically “homomorphic” sex chromosomes in various vertebrates (Nishida-Umehara et al., 1999; Schempp and Schmid, 1981). Heteromorphic sex chromosomes evolved from a pair of autosomes by a combination of suppression of recombination and accumulation of sexual antagonist genes (Ohno, 1967). The isolation of one of the sex chromosomes in one sex (Y chromosomes in mammals and some fish, the W chromosome in birds and many non-mammal vertebrates) has led to degeneration and massive gene loss. The evolution of heteromorphic sex chromosomes has been indicated to lead to a gene dosage difference between the sexes. In mammals this has resulted in the inactivation of one of the X chromosomes in female somatic cells.

X chromosome inactivation is a unique example where the status of chromatin can be changed from active to inactive (facultative heterochromatin) on a chromosome-wide level. In therian female mammals (marsupials and placental mammals), one of the X chromosomes in somatic cells is heterochromatic and late replicating (Holmquist, 1987; Lyon, 1961; Ohno et al., 1963; Schweizer et al., 1987; Takagi, 1974). This transcriptionally silenced and condensed X-chromosome is visible as a Barr body in somatic cells. In the third major group of mammals, the egg laying monotremes (platypuses and echidnas), it is less clear if X inactivation and late replication occurs. Earlier replication banding did not reveal obvious asynchronously replicating X chromosomes (Wrigley and Graves, 1988). More recently molecular cytological data suggests the platypus X-chromosomes display partial and gene specific forms of inactivation, but still undergo some level of asynchronous replication of X-specific genes (Deakin et al., 2008a; Ho et al., 2009). Furthermore, a wholesale shift in replication timing for the avian Z-chromosome, which shares extensive homology with the extraordinary ten sex chromosome system in monotremes, is not observed in male homogametic birds, indicating that this process is only present in therian mammals (Arnold et al., 2008; Grutzner et al., 2004; Rens et al., 2007; Veyrunes et al., 2008).

4.1 Chromatin marks behind X-inactivation

The X-inactivation process results in monoallelic expression of the vast majority of X-linked genes in humans and mice. Its process is dependent on critical elements which reside in the X-inactivation centre (XIC) on each X-chromosome, particularly the imprinted Xist and Tsix genes, and long-range chromatin elements (Boumil and Lee, 2001; Brockdorff et al., 1991; Brown et al., 1991; Clerc and Avner, 2003). The Tsix gene appears to regulate chromatin
structure at the *Xist* locus, causing its expression to be upregulated. This upregulation of *Xist* RNA corresponds to chromatin changes in the inactive X, most of which are associated with silencing (Heard, 2005). These *Xist*-induced marks on the inactive X include methylation of CpG dinucleotides in gene promoters, and histone modifications such as hypomethylation of H3K4 and hypoacetylation of H3K9 and H4, also monomethylation of H4K20 and trimethylation of H3K27, and finally H2AK119 ubiquitination (reviewed in Zakharova et al., 2009). Furthermore, the chromatin from the inactivated-X chromosome is enriched for the histone variant macroH2A1, and the final epigenetic mark is the late replication status of the inactive-X during the S-phase (reviewed in Zakharova et al., 2009). This inactivated state facilitates a change in the expression potential of the inactive X, and thus provides gene dosage compensation in female therian mammals (Hellman and Chess, 2007). It has also been observed that the active human X-chromosome is hypomethylated at gene-rich areas compared to the inactive X-chromosome, which displays hypermethylation (Hellman and Chess, 2007).

In placental mammals X inactivation of the maternal or paternal X chromosome is random, in marsupials and mouse extra-embryonic tissues only the paternal X is inactivated (reviewed in Lee, 2003). The epigenetic marks associated with marsupial X-inactivation include the loss or reduction of active histone marks on the inactive-X including H3K4 dimethylation, H4 acetylation, H3K9 acetylation marks (Koina et al., 2009; Wakefield et al., 1997). However, the absence of inactivating histone marks in marsupials, as observed on the inactive-X in placental mammals, may be due to the absence of a XIC region in marsupials (Duret et al., 2006; Hore et al., 2007; Koina et al., 2009). The evolution of the *Xist* non-coding RNA gene involves the pseudogenization of a protein-coding gene in the placental mammalian genome. As such, this gene is not present in marsupial and monotreme mammals, and cannot be found in the regions orthologous to the XIC in these mammalian clades. In marsupials and monotremes, the orthologous flanking genes to the placental mammal XIC region map to different ends of the X-chromosome and chromosome 6 respectively (Davidow et al., 2007; Deakin et al., 2008b; Duret et al., 2006; Hore et al., 2007; Shevchenko et al., 2007).

The FISH based dot assay was utilized to measure replication timing of genes from X-specific regions within the five platypus X-chromosomes. This did not reveal a clear cut replication asynchrony on X specific regions but one of the homologous pairs, namely the X3 chromosomes, showed significantly differential condensation, indicative of wholesale chromatin silencing (Ho et al., 2009). The other four sex chromosome pairs in platypus females, however, show no significant difference in condensation between homologs indicating that the X-inactivation process in monotremes may be region specific (Ho et al., 2009). In male homogametic birds (with ZZ sex chromosomes), studies have shown that whilst the entire chicken Z-chromosome replicates synchronously, the inactivation process appears to be partial and gene-specific, with dosage-compensation occurring stochastically, and in a stage and tissue-specific manner (Arnold et al., 2008; Deakin et al., 2008a; Ho et al., 2009; Kuroda et al., 2001; Kuroiwa et al., 2002; Mank and Ellegren, 2009). Moreover, there is evidence that dosage compensation in monotreme mammals operates in a similar manner as in birds, with platypus females showing stochastic transcriptional inhibition of genes from X-chromosomes (Deakin et al., 2008a). In this case, some X-genes were shown not to be dosage compensated, whilst monoallelic expression was observed at other X-chromosome loci (Deakin et al., 2008a).
5. Asynchronous replication in genes subject to genomic imprinting and allelic exclusion

Genomic imprinting refers to the parent of origin dependent monoallelic expression of an autosomal gene, engendered by the inheritance of parental-specific methylation at an allele. To date, imprinting mechanisms have only been found in therian mammals, which rely on extensive intrauterine foetal-maternal exchange during early development. The ‘parental conflict hypothesis’ proposed that imprinting is a way of parental genomes counteracting the effects of each other during foetal development, particularly in foetal-maternal placental nutrient exchange (Moore and Haig, 1991). Monotremes, unlike therian mammals, have a brief intrauterine foetal-maternal exchange and there is no competition of the parental genomes over maternal resources. In line with the ‘parental conflict hypothesis’ to date no imprinting has been discovered in this basal mammalian lineage, suggesting that imprinting evolved after their divergence from therian mammals (Renfree et al., 2009).

5.1 Imprinted genes

Imprinted genes are asynchronously replicated (Table 1), where the replication of one allele lags behind the other in the S-phase, even though the two alleles should be controlled by similarly situated ORs. Traditionally, imprinting involves DNA methylation at only one allele of a gene (i.e. the copy from just one parent is methylated) (Delcuve et al., 2009). In most cases the imprinted allele is methylated and transcriptionally silent. The active or silenced transcriptional state of an allele appears to go hand in hand with replication timing, whereby the expressed allele is early replicated, whilst the silenced allele undergoes late replication in the S-phase (reviewed in Zakharova et al., 2009).

Imprinting control regions (ICRs) are the elements which control the imprinting status of an allele (Bartolomei, 2009). The parentally inherited methylation status, which is established during gametogenesis, of an ICR dictates its control over an allele, meaning that maternal and paternal ICRs at a locus will interact differently with transcriptional control elements, due to their dissimilar methylation status (Bartolomei, 2009). Notably, maternally-imprinted ICRs are often found in the promoters for antisense transcripts, whilst paternally-imprinted ICRs usually reside in intergenic regions (reviewed in Edwards and Ferguson-Smith, 2007). Moreover, the formation of large imprinted gene clusters, where regions of maternally and paternally expressed genes are interspersed with non-imprinted genes, allows many imprinted genes to share regulatory elements, such as ICRs (reviewed in Bartolomei, 2009).

The asynchronously replicating status of imprinted loci has been linked to DNA methylation and other epigenetic marks associated with imprinted gene silencing (Dünzinger et al., 2005). However in birds, which have no fetal-maternal exchange and display no form of genomic imprinting, there are several conserved regions of mammalian imprinted gene orthologs that are asynchronously replicated (Dünzinger et al., 2005). These asynchronously replicating regions are found on chicken macrochromosomes which, compared to their microchromosome counterparts, are hypoacetylated, hypomethylated, late replicating, and display a lower recombination rate during meiosis (Consortium, 2004; Grutzner et al., 2001; McQueen et al., 1998; Schmid et al., 1989). This indicates that asynchronous replication predates imprinting, and that the common vertebrate ancestor of mammals and birds had genomic regions with a ‘pre-imprinted’ status, whereby asynchronous replication still occurred (Dünzinger et al., 2005). It will be interesting to see whether monotreme orthologs of imprinted genes also replicate asynchronously, as observed in birds (Dünzinger et al., 2005).
5.2 Allelic exclusion genes

Allelic exclusion is a process whereby the future expression from one allele of a locus is chosen in a cell, resulting in monoallelic expression at the locus. Allelic exclusion is a feature of many multigene families, with olfactory gene clusters and immunoglobulin gene clusters being two classic groups of genes utilizing this form of epigenetic control. However there are also other groups of genes which utilize allelic exclusion, including interleukins and the p120 catenin (Gimelbrant et al., 2005; Hollander et al., 1998). Many epigenetic elements control the cell’s choice over which allele will be active, including cis and trans-acting DNA sequences, long-range interactions, and chromatin modification (reviewed in Zakharova et al., 2009).

5.2.1 Olfactory genes

Whilst some olfactory receptor (ORc) genes are dispersed in the mammalian genome, many exist in clusters (Kambere and Lane, 2007). The largest cluster in mouse consists of 244 ORc genes, whilst in human the largest cluster contains 116 genes (Godfrey et al., 2004; Malnic et al., 2004). Both species have individual ORc genes and ORc clusters spread across many different chromosomes, with a few chromosomes containing large clusters of ORcs (Glusman et al., 2001; Kambere and Lane, 2007). However, even though the eutherian genome contains around 1000 ORc genes, only a single ORc gene will be expressed in a single olfactory neuron, meaning that that neuron will only express one type of odorant receptor (Malnic et al., 1999). Furthermore in a process known as allelic inactivation, the locus that is being expressed undergoes differential epigenetic processes at each allele that cause one allele to be inactivated, and thus monoallelic expression of the gene (Chess et al., 1994).

Chromosome conformation capture (3C) assays have given an insight into the mechanisms surrounding the selection of a single ORc gene (Lomvardas et al., 2006; Serizawa et al., 2003). The recently developed 3C technique has become invaluable to studies on nuclear architecture, as it is able to detect and quantify long-range DNA interactions in vivo, at high resolution, between sequences in close nuclear proximity. The technique relies on the cross-linking of proteins using formaldehyde in intact nuclei or cells (Dekker et al., 2002). The result is that proteins are cross-linked to other proteins and to adjacent chromatin (Orlando et al., 1997). DNA regions that are actually touching at the time of fixation will be held together via the cross-linking of their DNA bound proteins. The cross-linked genomic DNA is then digested with DNA restriction enzymes and the resulting DNA segments are then ligated. Finally, PCR across these ligation sites detects long-range interacting regions at the DNA sequence level (Dekker et al., 2002).

The 3C experiments on olfactory neurons indicated that ORCs undergo an interaction with a long-range interacting region called the “H element”, located within the mouse ORc gene cluster MOR28, and perhaps do so in a competitive manner in order to become the activated ORc gene (Fuss et al., 2007; Lomvardas et al., 2006; Serizawa et al., 2003) so that only one gene will be chosen and actively expressed (Lomvardas et al., 2006; Serizawa et al., 2003). However another study showed that deletion of the H element only affected proximal genes within its MOR28 cluster, with no effect on genes outside this cluster, indicating that it cannot be the only factor involved in terms of activating ORc genes in long-range cis and trans conformations (Fuss et al., 2007).
ORc genes are observed to undergo asynchronous replication (Table 1), with different clusters and individual ORc genes on the same chromosome undergoing replication at the same time in the S-phase, and the establishment of this form of replication occurring in early embryogenesis (Chess et al., 1994; Mostoslavsky et al., 2001; Singh et al., 2003). The asynchronous replication of ORc loci is believed to be controlled in part by the Polycomb group methyltransferase Eed, as ORc genes lose their asynchronously replicating status in its absence (Alexander et al., 2007). This could explain how ORc genes located on the same chromosome are observed to undergo asynchronous replication, with Eed being a requirement for asynchronous replication, regardless of position on a chromosome (Alexander et al., 2007; Singh et al., 2003).

5.2.2 Immunoglobulin gene loci

It has been suggested that asynchronous replication plays an important role in the selection of which parental allele will undergo V(D)J rearrangement. The allelic exclusion process in mouse occurs for the genes which do not undergo intrachromosomal recombination, and thus are silenced. The rearrangement process of the immunoglobulin genes in mouse requires crosstalk between two loci from two different chromosomes, namely the IgH locus (containing V, D and J gene segments), and Igκ locus (containing V and J segments). The de novo methylation of all the VDJ alleles occurs at the implantation stage, and this is also when asynchronous replication is established (Table 1) (Mostoslavsky et al., 2001). However, the selection of one allele at each locus to undergo early replication puts this allele down a demethylation and chromatin opening pathway, allowing it to be rearranged and to become a functional gene (Goldmit et al., 2002). The other late replicating allele however, remains methylated and cannot be rearranged, and is therefore functionally silenced (Goldmit et al., 2002). The two alleles also have different histone marks with the inactive allele binding the heterochromatin specific protein HP1, and the active allele displaying active histone marks such as di- or trimethylated H3K4, and H3 and H4 acetylation (reviewed in Zakharova et al., 2009).

Asynchronous replication and monoallelic expression are hallmarks of genes which undergo imprinting, X-inactivation, and allelic exclusion. Whilst each might come with its own epigenetic makeup, there are also similarities in the types of epigenetic marks observed to differentiate the active allele (with active histone marks) from the inactive allele (with silencing histone marks). Furthermore, the very fact that asynchronous replication occurs together with different forms of epigenetic monoallelic expression suggests that asynchronous replication may have evolved as a mechanism to control the expression of underlying genes, helping to establish the correct epigenetic marks for monoallelic expression.

6. The CTCF protein and the interactome

The CCCTC-binding factor (CTCF) is a renowned genome organiser, and has roles in regulating long-range chromatin interactions (both intrachromosomal and interchromosomal), but also has roles in other processes such as transcriptional insulation, activation/repression, imprinting control, and X-inactivation (Ling et al., 2006; Murrell et al., 2004; Phillips and Corces, 2009). It is also implicated to have roles in sister chromatid cohesion during DNA replication, as CTCF has been shown to interact with the STAG1
<table>
<thead>
<tr>
<th>Gene/Region</th>
<th>Placentals</th>
<th>Monotremes (Platypus)</th>
<th>Birds (Chicken)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex chromosone specific regions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xist</td>
<td>39%</td>
<td>Mus 138</td>
<td>(Gribnau et al., 2005)</td>
</tr>
<tr>
<td>Mecp2</td>
<td>33%</td>
<td>Mus 108</td>
<td>(Gribnau et al., 2005)</td>
</tr>
<tr>
<td>Smcx</td>
<td>38%</td>
<td>Mus 157</td>
<td>(Gribnau et al., 2005)</td>
</tr>
<tr>
<td>OGN</td>
<td>NA</td>
<td>22% 587</td>
<td>(Ho et al., 2009)</td>
</tr>
<tr>
<td>APC</td>
<td></td>
<td>29% 420</td>
<td>(Ho et al., 2009)</td>
</tr>
<tr>
<td>Imprinted genes (in eutherians)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Igf2</td>
<td>23%</td>
<td>Mus &gt;100</td>
<td>(Kitsberg et al., 1993)</td>
</tr>
<tr>
<td>Igf2R</td>
<td>35%</td>
<td>Mus &gt;100</td>
<td>(Kitsberg et al., 1993)</td>
</tr>
<tr>
<td>Mest/Co pg2</td>
<td>25%</td>
<td>HSA &gt;200</td>
<td>(Bentley et al., 2003)</td>
</tr>
<tr>
<td>TCRβ</td>
<td>46%</td>
<td>Mus 100-300</td>
<td>(Mostoslavsky et al., 2001)</td>
</tr>
<tr>
<td>B-cell receptor (κ)</td>
<td>48%</td>
<td>Mus 100-300</td>
<td>(Mostoslavsky et al., 2001)</td>
</tr>
<tr>
<td>IL-2</td>
<td>68%</td>
<td>Mus 100</td>
<td>(Hollander et al., 1998)</td>
</tr>
<tr>
<td>Olfactory receptor</td>
<td>31%</td>
<td>Mus &gt;99</td>
<td>(Simon et al., 1999)</td>
</tr>
</tbody>
</table>

Table 1. A subset of the asynchronous replication data that exists for genes/genomic regions which are sex chromosome specific, imprinted (in eutherians), or undergo allelic exclusion.

subunit of cohesin and localize cohesin to specific CTCF binding sites on chromosome arms (Parelho et al., 2008; Rubio et al., 2008). Important in this context is that the CTCF protein has been shown to mediate asynchronous replication and imprinting control for the Igf2-H19 cluster (Bergstrom et al., 2007).

6.1 The evolution of CTCF
The CTCF protein is highly conserved across higher eukaryotes, and the active site shows close to 100% homology between mouse, human and chicken suggesting that the protein has a highly conserved role (Ohlsson et al., 2001). A CTCF gene duplication event is believed to have occurred in the amniote ancestor preceding the divergence of reptiles and birds, as they both have functional CTCF, but not its gene paralogue, BORIS (brother of regulator of imprinted sites) (Hore et al., 2008). BORIS has similar DNA binding capabilities to CTCF, but shows antagonistic epigenetic regulation to CTCF, as well as gonad-specific expression in placental and marsupial mammals (Hore et al., 2008). Conversely, BORIS appears to be widely expressed in monotremes and reptiles, indicating that the gene underwent a
functional change after the divergence of therian mammals, which is interesting because as yet, there is no evidence that CTCF binding sites exist in the genomes of earlier diverged monotreme mammals (Hore et al., 2008; Weidman et al., 2004). However, CTCF sites have been observed in the chicken genome which is an earlier-split vertebrate than the monotreme clade, and tied with the evidence that CTCF binding occurs in therian genomes (Banaihmad et al., 1990; Lobanenkov et al., 1990), it is likely that CTCF sites exist in the monotreme genome.

6.2 CTCF and genome organization

It is hypothesised that although chromatin fibres are subjected to random contacts, and thus will always inhabit slightly different positions in the nucleus, the characteristics of the interacting regions on chromosomes allow interactions to occur (de Laat and Grosveld, 2007). Furthermore, it has been argued that genomic regions preferentially interact with other genomic regions that have similar characteristics to their own, such as regions that share CTCF binding (de Laat and Grosveld, 2007). It has been hypothesised that regions of a chromosome which undergo similar replication timing, like asynchronously replicating genes, may be pulled into similar replication domains (Ryba et al., 2010; Singh et al., 2003). Within the mammalian cell nucleus, chromatin from separate chromosomes is organised into the aforementioned chromosome territories. Within these CTs, a higher order of chromatin structure exists, where domains containing specific chromosomal arms and bands have been found to be located in the nucleus in similar regions of certain cell types (Dietzel et al., 1998). Genes are readily transcribed when they reside on the periphery of chromosome territories, and can even loop out of the territories. Furthermore, genes that are late-replicating and inactivated are often seen to reside on the outer regions of chromosome territories near the nuclear periphery. Looping of the chromatin fibres allows genes to easily interact with the transcriptional machinery residing in the interchromatin compartments (Cremer and Cremer, 2001; Osborne et al., 2004). Imprinted and allelic exclusion genes often ‘loop out’ and undergo long-range interactions for regulatory purposes (Ling and Hoffman, 2007; Lomvardas et al., 2006).

A good example of CTCF controlling some of the discussed epigenetic, replication, and transcriptional mechanisms occurs at the imprinted Igf2/H19 domain. The ICR for this imprinted cluster lies between these two genes, in the 5’ flanking sequence of H19, and the maternal allele interacts with CTCF (Kurukutti et al., 2006). CTCF regulates and insulates imprinted gene transcription for the Igf2/H19 region by controlling the intrachromosomal interactions of the maternal and paternal alleles (Murrell et al., 2004). When endogenous CTCF is knocked-down in mice, loss of Igf2 imprinting is observed, whilst deletion of the ICR leads to biallelic expression of H19 (Ling et al., 2006). In mouse, the paternal chromosome forms a DNA loop between the differentially methylated region (DMR) 2, present in the Igf2 gene, and the methylated ICR, aided by putative binding factors (Murrell et al., 2004). When the paternal Igf2 allele promoter comes into close proximity with the H19 enhancer elements, Igf2 transcription occurs (Murrell et al., 2004). The DMR1 on the maternal chromosome interacts with the unmethylated ICR, which causes the maternal Igf2 allele to be sequestered into a transcriptional silencing loop. This causes the maternal H19 allele to become proximal to its enhancers, allowing it to be expressed (Murrell et al., 2004). Conversely, CTCF also facilitates an interchromosomal interaction in mouse between the Igf2/H19 domain, and the Wsb1/Nf1 region on a different chromosome (Ling et al., 2006).
Specifically, the ICR in the imprinted Igf2/H19 domain, which contains CTCF binding sites, has been found to interact with another region with CTCF binding sites between the Wsb1 (WD repeat and SOCS box-containing 1) and Nfl (Neurofibromin 1) genes (Ling et al., 2006). Whilst the Wsb1 and Nfl do not appear to be imprinted, as their expression is biallelic, only the paternal copy of the Wsb1/Nfl region interacts with CTCF (Krueger and Osborne, 2006; Ling et al., 2006). As explained before, CTCF only binds the maternal copy of the ICR region (flanked by Igf2 and H19). It is consequently hypothesized that the long-range interaction observed between the ICR and Wsb1/Nfl region occurs between the maternal and paternal copies respectively, and is mediated by the genome-organizing protein CTCF (Ling et al., 2006).

6.3 Replication timing and CTCF
The specific binding of CTCF at the maternal ICR in the mouse Igf2/H19 domain has been shown to mediate asynchronous replication in this imprinted region (Bergstrom et al., 2007). The inheritance of a mutated maternal ICR, which lacks CTCF binding, caused the usually late replicating maternal Igf2/H19 domain to become early replicating (Bergstrom et al., 2007) showing that CTCF binding is required for asynchronous replication of these loci. The mechanism by which CTCF might regulate asynchronous replication at this domain, however, is still unclear. In addition to replication CTCF is involved in other epigenetic effects, including long-range interactions (both intrachromosomal and interchromosomal), insulator activity and transcriptional activation (Kurukuti et al., 2006; Ohlsson et al., 2001; Zhao et al., 2006). Notably, it has been shown that regions which undergo greater amounts of long-range chromatin interaction are subject to late replication timing (Ryba et al., 2010). Another example of the close relationship between replication, CTCF, and methylation occurs at the differentially methylated silencer region controlling the expression of the AWT1/WT1-AS genes (Hancock et al., 2007). The CTCF protein can only bind the late-replicating unmethylated paternal silencer region within the AWT1/WT1-AS cluster, allowing expression of the paternal alleles. The homologous early-replicating maternal region however, has a methylated silencer which does not facilitate CTCF binding and so the maternal AWT1/WT1-AS alleles are not expressed (Hancock et al., 2007). It is interesting to speculate as to whether CTCF also controls the asynchronous replication observed at the WT1 locus in human, and perhaps even in birds (Bickmore and Carothers, 1995; Dünzinger et al., 2005). It is also interesting to note that in both cases the late-replicating allele at these imprinted loci, namely in the maternal Igf2/H19 allele and the paternal AWT1/WT1-AS allele, is the allele which binds CTCF (Bergstrom et al., 2007; Bickmore and Carothers, 1995; Hancock et al., 2007). Whilst CTCF is observed to mediate asynchronous replication and imprinting at the Igf2/H19 domain in eutherian mammals, the fact that the imprinted orthologs of Igf2/H19 and AWT1/WT1-AS still asynchronously replicate could suggest that CTCF binding in these regions evolved before establishment of genomic imprinting.

6.4 The role of CTCF in replication timing changes in cancer
CTCF may also play a role in the progression of cancer and has many of the characteristics of a tumour suppressor gene; in the human genome it maps to a small region, 16q22.1, which characteristically undergoes loss of heterozygosity in many solid tumours (reviewed in Filippova et al., 1998). Furthermore, changes in DNA consensus sites and DNA methylation patterns in cancers are known to cause loss of CTCF binding, which could result in the loss of functional control of these regions (Filippova et al., 2002; Ohlsson et al., 2001). The regions required for zinc-finger formation, and their corresponding DNA binding
sites are often mutated in tumours, changing the CTCF binding-landscape of a genome (Filippova et al., 2002). Specifically, the presence of these mutations in tumours was observed to abolish CTCF’s association with the Igf2/H19 growth regulating genes, whilst not changing its association with non-growth regulating genes (Filippova et al., 2002; Ohlsson et al., 2001). The loss of CTCF association with the Igf2/H19 region in tumours could be associated with a shift in replication asynchrony. As mentioned in the previous section, when CTCF binding is abolished in the maternal Igf2/H19 region it results in the loss of asynchronous replication at the locus (Bergstrom et al., 2007). Furthermore, omission of CTCF binding to the maternal Igf2/H19 ICR has also been observed to abrogate inter-chromosomal interactions for this region (Ling et al., 2006). These results all indicate that the loss of CTCF binding for specific genomic regions in tumours has downstream epigenetic effects, such as loss of replication asynchrony and chromatin interaction, for the genes usually involved in CTCF-interaction.

7. Evolution of replication timing and epigenetic control

7.1 The evolution of replication timing

At the genome level, recent work shows that asynchronous replication pre-dates the establishment of monoallelic expression and genomic imprinting (Zechner et al. 2006, Wright et al. in preparation). The bird genome, which lacks genomic imprinting, contains conserved regions of mammalian imprinted gene orthologs that are asynchronously replicated (Dünzinger et al., 2005). This indicates that asynchronous replication most likely pre-dates imprinting, and that the common vertebrate ancestor of mammals and birds had genomic regions with a ‘pre-imprinted’ status which still underwent asynchronous replication without any form of traditional imprinting (Dünzinger et al., 2005). It is interesting to note that a recent genome-wide study has indicated that regions with conserved synteny also have conserved replication profiles among human and mouse (e.g. Ryba et al., 2010). Imprinted clusters are renowned for having conserved synteny, and it has been suggested that the selection of highly conserved arrays of imprinted gene orthologs occurred during vertebrate evolution, however why these regions were selected for syntenic conservation has been difficult to explain (Dünzinger et al., 2005).

At the replicon level, there has been a model proposing that spatiotemporal properties of mammalian ORs contribute to a combination of pre-determined and stochastic DNA replication (Takahashi, 1987). This mechanism is echoed in budding yeast, which also shows OR activation in a combined chronological and stochastic manner (Barberis et al., 2010; Spiesser et al., 2009). This model, combined with the finding that conserved syntenic regions in human and mouse have very similar replication profiles, indicates that there is a conservation of the temporal programme controlling replicon firing. Furthermore there appears to be a highly conserved order in which amniote imprinted genes or imprinted gene orthologs replicate; with individual imprinted genes following similar temporal patterns when entering replication in birds, monotremes, and eutherians (Wright et al. in preparation). This indicates that in closer related clades of eukaryotes, this temporal replication program may be highly conserved.

7.2 The chromatin interactome and replication profiling

Developing molecular technologies are allowing greater insights into the many interactions occurring in a genome, but also showing how spatial organisation can affect other processes
in a genome, such as replication timing. Extensions of the previously discussed 3C molecular interaction technology include Associative Chromosome Trap (ACT), Circular Chromosome Conformation Capture or Chromosome Conformation Capture-on-Chip (4C), and Carbon-Copy Chromosome Conformation Capture (5C), all of which can measure more than a single to single region interaction (Dekker et al., 2002; Dostie et al., 2006; Ling et al., 2006; Simonis et al., 2006; Zhao et al., 2006). In addition to these technologies, new techniques are allowing interactions to be measured across entire genomes, resulting in the mapping of an “interactome”, whereby all the long-range interactions occurring in a genome are measured (Fullwood et al., 2009; Lieberman-Aiden et al., 2009). Specifically, there are two techniques that have been developed to do this, Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET) and Hi-C (which measures the three-dimensional architecture of a genome by coupling proximity-based ligation with parallel sequencing) (Fullwood et al., 2009; Lieberman-Aiden et al., 2009). These experiments, in conjunction with replication-timing profiling by microarrays, have indicated that the interactome of a genome is very closely aligned with replication timing (Ryba et al., 2010). The chromatin “interactome” is now understood to play a critical part in genome organisation; allowing complex regulatory networks of interactions to occur, each of which with functional significance, all of which highly dynamic and organised within a nucleus by proteins such as CTCF and the Estrogen-receptor alpha (Botta et al., 2010; Fullwood et al., 2009). These interactions also appear to be conserved in similar cell types across mammalian evolution, suggesting that perhaps these long-range interactions are part of an evolutionary conserved mechanism of spatial organisation (Ryba et al., 2010). Furthermore, initiation of replication appears to be an evolutionarily conserved process across eukaryotic evolution, and the overlay of entire genome replication timing profiles with interactome maps have shown that late-replicating regions are often undergoing greater amounts of long-range interaction (Ryba et al., 2010). These findings, in conjunction with asynchronous replication data, could indicate that long-range interactions which occur in abundance at imprinted and monoallelically expressed loci, are affecting asynchronous replication. Specifically, there is data supporting the argument that the allele undergoing long-range interaction could also be the allele which undergoes late-replication. Firstly, it has been observed that asynchronously replicated alleles often localize to spatially distinct regions in a nucleus (Gribnau et al., 2003; Sadoni et al., 1999). Secondly, as mentioned previously, the late-replicating maternal \( Igf2/H19 \) allele and the paternal \( AWT1/WT1-AS \) allele, are also the alleles which bind CTCF, in an imprinting dependent manner. It could be that the binding of proteins which mediate long-range chromatin interaction at these alleles is facilitating greater amounts of interaction, which is reflected in their late replicating status, and also in the asynchronous replication of these genes (Bergstrom et al., 2007; Bickmore and Carothers, 1995; Hancock et al., 2007).

7.3 Measuring replication to combat cancer

It has been proposed that measuring changes in replication profiles may be a way of detecting abnormalities associated with cancer, not observed through usual techniques (reviewed in Watanabe and Maekawa, 2010). Epigenetic reprogramming in diseased cells is often observed to occur with changes in replication timing patterns, with changes in replication being observed with chromosomal rearrangements in cancer cell lines (D’Antoni et al., 2004; Gondor and Ohlsson, 2009; State et al., 2003). Better detection of prostate cancer
may come in the form of measuring replication timing changes observed in peripheral blood lymphocytes undergoing aneuploidy (Dotan et al., 2004). In terms of protein detection of cancer, measuring the function of the tumour suppressor gene p53, may be a good determinant in the progression of cancer. P53 is the most commonly mutated gene in human cancers, and is a G1/S-phase and S-phase checkpoint regulator during DNA replication. Loss of its function is observed to affect the replication timing of human colon carcinoma cells (Watanabe et al., 2007).

Changes in replication timing may also be affected by altered function of CTCF in cancer. As mentioned previously, it has been observed that mutation of CTCF binding sites near growth factor genes, such as in the Igf2/H19 region, occurs in many tumours (Filippova et al., 2002). These mutations may cause a loss of CTCF binding in the region, which has been observed to abolish asynchronous replication of the Igf2 locus, and changes the replication timing of the gene (Bergstrom et al., 2007). However the mutation of CTCF binding sites would also change the interactome profile of a cell. Loss of CTCF-binding through mutation around genes like Igf2 and H19 would result in them no longer undergoing their “normal” chromatin interactions, perhaps causing different spatial organization of these loci in the nucleus of a cancerous cell.

7.4 The chromatin interactome: controlling eukaryotic replication timing
To date there is a lack of data that could provide insight about the evolution of an interactome. It has been observed that many long-range interacting regions share many of the same (but not necessarily all) epigenetic characteristics, such as asynchronous replication, monoallelic expression, differentially methylated regions and histone modifications and variants, imprinting, and CTCF binding. It is currently unknown how these epigenetic events evolved and investigating those epigenetic features in a range of vertebrate genomes could tease apart the sequence of events that has led to a complex network of epigenetic regulation.

Chromatin interactions may have evolved in many genomic control processes, but it is the binding of master genome regulators, like CTCF, which dictate where these interactions can occur. The CTCF protein is highly conserved among amniotes, conserved in vertebrates, and exists in Drosophila and subsets of nematodes (Heger et al., 2009; Ohlsson et al., 2001). Furthermore, there is evidence to suggest that CTCF binding and function are conserved in humans, mouse, and chicken, in genes such as β-globin, whereby CTCF binding at this locus allows cell-type specific intrachromosomal interactions to occur (Bell et al., 1999; Yusufzai et al., 2004). CTCF binding and chromatin interaction in this region suggest that CTCF spatial control of chromatin, at least in this region, was present in the common ancestor of amniotes. The evolutionary conservation of replication timing and the strikingly similar genomic interactome in similar cell types among human and mouse suggests that replication timing is intrinsically tied to long-range interaction. Moreover, there is evidence to suggest that replication timing relies on the presence of long-range interactions at specific loci, with the knockdown of long-range mediator proteins causing interactions to be abolished, and also causing replication asynchrony to cease (Bergstrom et al., 2007; Fullwood et al., 2009; Ling et al., 2006). The loss of replication asynchrony in this case could be due to ectopic spatial organisation of the alleles, whereby the loss of the interaction mediator protein causes the allele of a locus to reside in an atypical subnuclear domain. This irregular replication domain would not have the correct molecular and chemical
characteristics to allow the ORs of the spatially ectopic allele to fire in the normal temporal order. This could cause the erroneous firing of ORs in such a way as to abolish replication asynchrony at the locus.

8. Conclusion

Replication timing of DNA at S-phase is tightly regulated and affects gene activity, nuclear organisation, as well as other aspects of genome biology. Differences in replication timing have been used to identify individual chromosomes and differentiated sex chromosomes for several decades. Since then, an increasing number of proteins have been identified as important for regulating replication timing and genome-wide approaches are now used to study replication timing. A fascinating variation of the replication-timing theme is asynchronous replication, which appears to be closely aligned with other epigenetic mechanisms involved in long-range interaction, genomic imprinting and X chromosome inactivation. Whilst previous research has stipulated that asynchronous replication and long range interactions have evolved as a result of epigenetic control of (eg. monoallelic expression), there is emerging evidence that both predate the presence of other epigenetic processes. We suggest that the interactome has played a role in the evolution of spatial nuclear organisation. In addition, mutations in sequences important for long-range interaction and replication timing, and also changes in the replication timing program itself, are important factors influencing a diverse array of human diseases, including cancer. The study of replication timing in different organisms and in human disease will reveal the full extent to which replication timing contributes to the epigenetic landscape in normal and abnormal cells.

9. Acknowledgements

Megan L. Wright is funded by an Australian Postgraduate Award. Frank Grutzner is an ARC Senior Research Fellow.

10. References


The study of DNA advanced human knowledge in a way comparable to the major theories in physics, surpassed only by discoveries such as fire or the number zero. However, it also created conceptual shortcuts, beliefs and misunderstandings that obscure the natural phenomena, hindering its better understanding. The deep conviction that no human knowledge is perfect, but only perfectible, should function as a fair safeguard against scientific dogmatism and enable open discussion. With this aim, this book will offer to its readers 30 chapters on current trends in the field of DNA replication. As several contributions in this book show, the study of DNA will continue for a while to be a leading front of scientific activities.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: