The human protamine gene cluster consists of three tightly regulated genes, protamine 1 (PRM1), protamine 2 (PRM2) and transition protein 2 (TNP2). Their products are required to repackage the paternal genome during spermiogenesis into a functional gamete. They reside within a single DNase I-sensitive domain associated with the sperm nuclear matrix, bounded by two haploid-specific Matrix Attachment Regions. The nuclear matrix is a dynamic proteinaceous network that is associated with both transcription and replication. While substantial effort has been directed toward pre- and post-transcriptional regulation, the role of the nuclear matrix in regulating haploid expressed genes has received comparatively little attention. In this regard, the functional organization of the human PRM1→PRM2→TNP2 cluster and where appropriate, comparisons to other model systems will be considered.

Introduction

We are continuing the exploration of the nucleus that commenced in 1871 when Miescher began exploiting the gonad of the spawning male salmon in the Rhine River as an alternative source of cells from which to investigate nuclear composition. A mixture of protamine encased nucleic acids termed “nuclein” was isolated (Miescher 1874). Once again attention is being focused toward understanding the carefully ordered structure of the cell nucleus. It is becoming apparent that the manner in which chromatin is organized within the nucleus provides a door to understanding gene regulation and cellular reprogramming.

Individual chromosomes occupy distinct territories within the cell nucleus. Transcriptionally active segments tend to localize to the periphery of the territories whereas transcriptionally inert regions localize to their centers (Kurz et al. 1996). In proliferating cells, their central position within the nucleus positively correlates with gene density (Boyle et al. 2001). Changes in the relative position can reflect the cell entering different stages of differentiation (Foster et al. 2005), undergoing malignant transformation (Cremer et al. 2003) or senescence (Bridger et al. 2000). Key to this organization (Ma et al. 1999) is the attachment or association of the genome to a network of proteins that lies just interior to the nuclear envelope, termed the nuclear matrix (Ma et al. 1999).

Until recently, the nuclear matrix had been regarded as a static structure, arranging chromatin into domains varying in length from ~100 - 200 kb in somatic cells and 20 - 50 kb in sperm (Barone et al. 1994). The points of attachment, i.e., the MARs, Matrix Attachment Regions, vary in size from 100 - 1000 base pairs. Several “families” of MAR motifs that anchor chroma-

Corresponding author E-mail: steve@compbio.med.wayne.edu
tin to the nuclear matrix or recruit specific trans-factors have been described (reviewed in Platts et al. 2006). It is slowly becoming accepted that nuclear matrix association is at the heart of a series of nuclear events (Kramer et al. 1997; Ostermeier et al. 2003; Vassetzzy et al. 2000; Yasui et al. 2002; Zaidi et al. 2005). In some cases they may shield domains from neighbouring enhancers, the silencing effects of heterochromatin (Martins et al. 2004, Namciu et al. 1998) and may even be required for early embryonic development (Ward et al. 1999). Structural elements of the nuclear matrix have been shown to recruit a number of chromatin modifiers (Yasui et al. 2002). Their detailed molecular interacations of association and mechanisms of action are only beginning to be characterized (Heng et al. 2004).

Factors associated with many of the underlying nuclear processes are often co-purified with nuclear matrix proteins. One well characterized family is that of the Special AT-rich sequence Binding proteins SatB1 and SatB2. These MAR-binding proteins appear as cage-like networks (Cai et al. 2003) providing a MAR associating platform for specific loci throughout the nuclear interior. They modulate differentiation of several cell lineages acting as both activators and silencers likely reflecting their phosphorylation status (Pavan Kumar et al. 2006).

SatB1 coordinates the interaction of several factors critical for T-cell differentiation (Alvarez et al. 2000). For example, silencing is achieved through site-specific recruitment of histone deacetylases and ATP-dependent nucleosome remodeling complexes (Cai et al. 2003; Yasui et al. 2002) to the IL-2Ra locus MAR. SatB1 can also promote transcription of epsilon-globin in erythroid progenitors through its interaction with CBP (Wen et al. 2005). Similarly, craniofacial patterning and bone formation are modulated through the interaction of SatB2 with several targets. For example, interaction of SatB2 with an enhancer downstream of Hoxa2 modulates skeletal patterning whereas interaction with transcription factors Runx2 and AFT4, synergistically acts to drive osteoblast differentiation (Dobreva et al. 2006). Like other MAR-binding proteins, members of the Runx family are characterized by their nuclear matrix targeting signal. Visualization at interphase reveals punctuate scaffolds. Perhaps these scaffolds are preserved during mitosis to promote the next active phase of transcription (reviewed in Zaidi et al. 2005). Epigenetic regulation likely reflects nuclear matrix association.

Developmental systems including spermatogenesis have been used as models to study epigenetic regulation (Kramer et al. 1998). Encased within the walls of the seminiferous tubules, each phase of this continuous process is highly ordered. The pathway is marked by a series of morphologically distinct transitions that ultimately yield a highly compacted haploid genome encased in a unique motile delivery vehicle. For example, the paternal genome is repackaged and compacted to approximately 1/13th the volume of the oocyte nucleus. Yet, it contains a complete set of epigenetic instructions that provide a key to the correct usage of the paternal genome. An overview of the utility of this model towards understanding how nuclear organization modulates the molecular mechanisms in differentiation of the PRM1→PRM2→TNP2 domain is presented.

Spermatogenesis

Spermatogenesis is characterized by a wave-like continuum of cellular and sub-cellular morphological, biochemical and physiological changes culminating in the repackaging of the male haploid genome (Sassone-Corsi 2002). The process initiates by the Bone Morphogenetic Protein-signaling of the proximal epiblast (reviewed in Raz 2005) giving rise to the primordial germ cells. The mitotically arrested germ cells then migrate to the genital ridge to form the presumptive gonad. During their migration they undergo epigenetic germline reprogramming, whereby the majority of the epigenetic marks, including imprints, are erased. The marks are
Repackaging the sperm genome

3

re-established during spermatogenesis prior to the paternal genome assuming a hypermethylated state. Interestingly, even though the PRM1→PRM2→TNP2 domain is hypermethylated, it is one of the few loci that remains in a potentiated chromatin conformation (Schmid et al. 2001). Following fertilization, the paternal genome is demethylated, save for the paternal imprints (Olek & Walter 1997). The underlying principle governing this state of epigenetic flux is uncertain. Together, demethylation along with the preservation of the paternal imprint likely impacts totipotency.

As summarized in Fig. 1, spermatogenesis can be divided into three phases 1) mitotic renewal of the spermatogonial stem cells. 2) Commitment to a meiotic phase reducing tetraploid spermatocytes to haploid round spermatids. 3). The final phase, spermiogenesis, the morphogenic differentiation of round spermatids to spermatozoa. Spermiogenesis then follows during which spermatozoa become motile. Most stages are easily identified reflecting the coordinated and gradual transition from a nucleo-histone organization to one that is dominated by nucleo-protamines. The morphologically distinct cell types represented within the cross section of the tubule shown in Fig. 1 can routinely be recovered by various cell-separation techniques. Spermatogonia are localized to the basal compartment, whereas the meiotic stages localize to the adluminal compartment. The differentiative process culminates in spermatozoa being released into the lumen where they collect and mature in the epididymus. This summarizes the continuous spermatogenic wave that in humans takes approximately 60 days to complete (Clermont 1972). Cells from the majority of stages can be enriched by their selection during the first wave and easily isolated in relatively pure form (Bellve 1993). This provides an ideal system to study differentiation.

Nuclear Organization during Spermatogenesis

Throughout the mitotic and meiotic programs, the germ cell nucleus is structured in a manner similar to that of somatic cells. The re-establishment of paternal imprints by targeted genomic methylation and histone acetylation continues through mitotic and meiotic divisions (reviewed in Rousseaux et al. 2005).

As shown in Fig. 2, nucleo-histone organization persists throughout the spermatocyte to just prior to spermiogenesis when the transition from histone to protamine packaging begins in earnest. During this early stage of spermatogenesis the nuclear DNA is organized as nucleosomes. These are comprised of 146 bp supercoiled repeated segments of DNA, that wrap 1.75 times around a histone octamer composed of two heterodimers of each of H2A-H2B plus H3-H4. Each is joined by approximately 15 bp of "linker DNA," stabilized by histone H1. This repeated unit then coils to form a chromatin loop that associates with the nuclear matrix.

Spermatogonia harbor several germ cell-specific histone variants, i.e., TH2A, H2A.X, TH3, and H3.3A in addition to the somatic H2A and H3 counterparts. Variants of H2B including TH2B, H2B-RP and ssH2B, H3, including H3.3B and H3F3/B and H1 including H1t, H112 and HILS1 systematically begin to replace their somatic counterparts from the primary spermatocyte stage onward (Kimmins & Sassone-Corsi 2005). Several post-translational modifications are required to ensure histone/protamine exchange and to maintain the fidelity of the genetic material. These include ubiquitination, H4 hyperacetylation, phosphorylation and ADP-ribosylation. Ubiquitination and phosphorylation of histones have been associated with the initial histone to transition protein exchange. H2A/H2B ubiquitination has been implicated in sperm chromatin reorganization (Roest et al. 1996). It is likely that ubiquitination targets histones for degradation, signaling their replacement by the transition proteins TNP1, TNP2, TNP3, TNP4, (Wouters-Tyrou et al. 1998) as the cell differentiates until the latter phase of the
Spermatogonia differentiate into spermatocytes, then into round spermatids and finally into spermatozoa, traversing into the lumen for transport to the epididymis. Spermatogenesis consists of three parts: mitotic amplification of primitive spermatogonia; spermatocytic progression through meiosis and genetic recombination to culminate in haploid spermatids; and spermiogenesis, the morphological differentiation of the haploid male germ cell. Spermatogonia that differentiate to type B spermatogonia are committed to the spermatogenic differentiative pathway. In the progression from zygotene to pachytene spermatocytes, the reorganization of the male genome commences. Initially, most somatic histones are replaced with testis-specific histone variants (TH). During spermiogenesis, these are gradually displaced by the transition proteins (TP) and eventually by the protamines. Repackaging of the genome in this manner leads to a higher degree of compaction in sperm nucleus that finally yields the mature spermatozoa.

Fig. 1: Spermatogenesis. Spermatogenesis is a continual process throughout the seminiferous tubule. Cells that can be isolated in relatively pure-form by various simple differential sedimentation techniques are illustrated. Just inside basement membrane, spermatogonia (1) differentiate into spermatocytes (2), then into round spermatids (3) and finally into spermatozoa (4), that traverse into the lumen (L), for transport to the epididymis where they mature further and acquire motility.

MITOSIS * MEIOSIS → SPERMIIOGENESIS

HISTONES TH TP PROTAMINES

round spermatid stage (Fig. 1). Although acetylation of select residues in histone amino termini is associated with actively transcribing chromatin, hyperacetylation of H4 is thought to be a critical step towards histone replacement. Foci throughout late-stage spermatogenic nuclei exhibit high levels of H4 hyperacetylation. These mark sites for localized and intermediate replacement (Lahn et al. 2002) with the phosphorylated group of transition proteins. These are subsequently replaced with protamines.
Fig. 2: Organization of somatic chromatin. The basic structural unit of eukaryotic somatic chromatin is the nucleosome. Left panel, approximately 146 bp of DNA supercoils 1.75 times around a histone octamer core composed of two H2A-H2B heterodimers and two H3-H4 heterodimers. Successive nucleosomes are joined by approximately 15 bp of "linker" DNA stabilized by Histone H1. The approximate diameter of the nucleosome chain is 10 nm. This is often referred to as the 'beads on a string' conformation. Middle panel, these structures supercoil onto themselves forming a 30 nm fiber. Right panel, DNA is further organized into 100 - 200 kb loops by attachment to the nuclear matrix by matrix attachment regions, MARs.

Following meiosis there is a burst of transcription, including that arising from the PRM1 and PRM2 genes (Kleene et al. 1983), most of which are then stored as inactive mRNPs (reviewed in Kleene 2003). Spermatid transcription subsequently ceases during the elongation phase (Kierszenbaum & Tres 1975). During this phase, the host of post-meiotically transcribed genes are released from their quiescent inactive mRNP bound state then translated into functional proteins (Fajardo et al. 1997). These include PRM1 and PRM2 mRNAs that are suppressed until the late elongating spermatid stage (Balhorn et al. 1984).

In mice and men, PRM1 is synthesized as a mature 51 aa protein whereas PRM2 is synthesized as a 102 aa and 107 aa precursor protein in human and mouse, respectively. This precursor is cleaved into a number of different peptide products to yield a 57 aa mature protein in humans (Wouters-Tyrou et al. 1998). After this initial processing, both PRM1 and PRM2 are phosphorylated. This reduces their net positive charge, facilitating transport and TNP replacement (Dadoune 1995). At this juncture, the intermediary nucleo-proteins are replaced with the newly translated arginine- and cysteine-rich protamines (Dadoune 1995). The protamines successively dephosphorylate facilitating nucleo-protein replacement with protamines then condense as intrastrand disulfide bond formation continues during spermiogenesis (Chirat et al. 1993). This completes the transition to a highly compacted structure.

With the constant upheaval of the nucleus throughout spermatogenesis, repair mechanisms are required to ensure the fidelity of the genetic material. Although it is not yet fully understood, ADP-ribosylation may provide at least part of the solution. Could this be part of the DNA-damage-dependent repair mechanism or act as a signal for recombination/repair? Consistent with this notion, members of the ADP-ribosylation pathway increase during the later stages of spermatogenesis when nuclear reorganization is prominent and DNA strand breaks accumulate (Mosgoeller et al. 1996). This would include nucleo-protamine remodeling (Meyer-Ficca et al. 2005) and/or post-fertilization genome decondensation and remodeling (Mosgoeller et al. 1996). Perhaps this reflects the management strategy to ensure the integrity of the paternal genome.
As modeled in Fig. 3, tracts of polyarginine associate directly with the major groove of DNA (Hud et al. 1994). The smaller size and the highly basic nature of the protamines package the DNA into a series of \( \sim 60 \) bp parallel sheets. One protamine dimer can span two helical turns or approximately 22 bp of DNA (Vilfan et al. 2004). These are stabilized as toroid ring structures through the formation of a disulfide bonded network among the adjacent protamines (Brewer et al. 1999; Dadoune 1995; Kimmins & Sassone-Corsi 2005). The toroid can be visualized as successive helical coils that form a hexagonal lattice, stabilized by intermolecular disulfide bonds that then associate with the sperm nuclear matrix using a unique suite of MARs. A consensus model of the protamine DNA toroid has emerged (Brewer et al. 1999; Vilfan et al. 2004) effectively resolving the comparatively dense structure of the male gamete nucleus. While this is an effective model, one must consider that approximately 15% of human DNA remains histone-associated (Tanphaichitr et al. 1978), while less than 2% of mouse DNA remains histone-associated (Balhorn et al. 1984). This begs the question how are the nucleo-histone regions organized?

Fig. 3: Organization of sperm chromatin by protamine. During spermiogenesis the fundamental nucleosome structure is ostensively replaced by protamine, compacting the genome approximately six-fold. (a) As proposed (Vilfan et al. 2004), a single protamine dimer is associated within the major groove (b) of 2 turns of DNA or approximately 22 bp. (c) The cysteine-rich protamines form intermolecular disulfide bonds, stabilizing the homodimer. (d) Protamine bound DNA then assembles into a series of loops that are stabilized by disulfide bridges that in turn form a toroid. This hexagonal lattice binds to the sperm nuclear matrix at approximately 20 - 50 kb intervals. (e) Entry into the toroid is indicated as S and the last loop exits the toroid through E. This figure was adapted and is reproduced with permission from J. Biol. Chem. Vilfan ID, Conwell CC & Hud NV 2004 Formation of Native-like Mammalian Sperm Cell Chromatin with Folded Bull Protamine. 279:20088-20095 from copyright 2004 Macmillan Magazines Ltd, The American Society for Biochemistry and Molecular Biology.

Nuclear organization of the PRM1→PRM2→TNP2 domain

After protamine deposition, specific segments of the PRM1→PRM2→TNP2 locus remain histone
enriched (Wykes & Krawetz 2003a) as shown in Fig. 4. The histone bound segments include specific MAR containing regions like the 5' MAR as well as the promoter regions of the PRM1→PRM2→TNP2 domain. The enhancer-promoter regions are similar among the various members of both the human and mouse loci (Wykes & Krawetz 2003b). A series of DNase I hypersensitive sites have been identified flanking the ends of the PRM1→PRM2→TNP2 domain (Wykes & Krawetz 2004). These sites colocalize to a series of predicted transcription factor binding sites including a binding site for the testes specific activator SOX-5 (Denny et al. 1992). This resides just downstream of the 5' MAR (Wykes & Krawetz 2004). Additional hypersensitive sites reside within the promoter regions (Wykes & Krawetz 2003b). These have been mapped to within ~300 nucleotides of the respective transcription initiation start sites. This includes the A box shown in Table 1, which is the most conserved cis-element among the members. It is likely indicative of the rigorous and coordinated manner in which these genes are regulated as a requisite for the survival of the species. For example, disruption at any phase can lead to a host of morphologically defective sperm and infertility phenotypes. On the one hand, ectopic or over-expression of PRM1 prematurely condenses the nuclear DNA thus effectively halting differentiation (Lee et al. 1995). On the other hand, haplo-insufficient transgenic mice lacking either PRM1 or PRM2 produce morphologically abnormal infertile sperm (Cho et al. 2001). Although not as severe, disruption of TNP2 expression by targeted mutagenesis presents as teratozoospermia and reduced fertility (Zhao et al. 2001).

![Fig. 4: The human protamine domain.](image)

The single DNase-I sensitive human PRM1→PRM2→TNP2 domain is bounded by two sperm-specific matrix attachment regions. The mouse domain is home to a similar array of MARs and conserved regulatory regions. However, as shown in Fig. 5, the mouse Prm1→Prm2→Tnp2 cluster is substantially compressed when compared to its human counterpart. While the mouse
Table 1. Summary of the common A box element among the promoters of the human and mouse protamine clusters.

<table>
<thead>
<tr>
<th>Species</th>
<th>Gene</th>
<th>Distance from TATA (bp)</th>
<th>A box sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>PRM1</td>
<td>102</td>
<td>GGGCTGCCC</td>
</tr>
<tr>
<td>Mouse</td>
<td>Prm1</td>
<td>105</td>
<td>GCCCTGCCC</td>
</tr>
<tr>
<td>Human</td>
<td>PRM2</td>
<td>65</td>
<td>GTGCCGCCC</td>
</tr>
<tr>
<td>Mouse</td>
<td>Prm2</td>
<td>73</td>
<td>GTGCCGCCC</td>
</tr>
<tr>
<td>Human</td>
<td>TNP2</td>
<td>96</td>
<td>GAGCTGCCC</td>
</tr>
<tr>
<td>Mouse</td>
<td>Tnp2</td>
<td>80</td>
<td>AAGCTGCCC</td>
</tr>
</tbody>
</table>

protamine genes do not themselves display a large degree of sequence divergence away from their human counterparts, the expansion of the intragenic region in humans reflects the substantial accumulation of repetitive sequence elements (Nelson & Krawetz 1994). This likely occurred after the initial divergence of murines and primates, as the mouse domain is essentially devoid of repetitive elements. Interestingly, the human sperm nuclear Alu elements appear
preferentially bound to protamine. This represents approximately 100,000 sequence elements in the genome. In comparison the mouse LINE/L1 elements appear histone bound (Pittoggi et al. 2000). Additionally, sperm histone/protamine-bound centromeres (Wykes & Krawetz 2003a) localize to the nuclear matrix (Yaron et al. 1998) whereas histone-bound telomeres localize to the nuclear periphery (Pittoggi et al. 1999; Zalenskaya et al. 2000). The significance of these observations remains to be delineated since the relative proportion of histones in both species varies markedly. Differential packaging may present an immediate target to assure the silencing of promoters from male haploid expressed genes like those of the PRM1→PRM2→TNP2 domain. Alternatively it may template the genome for histone repackaging upon fertilization or enable specific zygotic transcription from the paternal contribution.

Interactions with the nuclear matrix orchestrate a host of nuclear events in somatic cells (Kramer et al. 1998, 2000; Ostermeier et al. 2003; Vassetzky et al. 2000; Yasui et al. 2002). The sperm nuclear matrix plays an important supportive and structural role. Its constituents include factors involved in spermatozoal capacitation (Calzada & Martinez 2002), factors required for early oocyte activation (Fujimoto et al. 2004) and factors that likely assist with the formation of the male pronucleus (Ward et al. 1999). The MARs of the male haploid expressed human protamine locus are situated at the boundaries of a DNase I-sensitive domain (Fig. 4; Kramer & Krawetz 1996). They have been shown to provide critical components towards assuring non-ectopic expression and minimizing position effects both in an animal models (Martins et al. 2004) and in humans (Kramer et al. 1997). Data from the transgenic mouse model of the human PRM1→PRM2→TNP2 domain has suggested that the upstream and downstream MARs synergistically act to shield this segment from its neighboring chromatin environment. The 5' MAR may act to enhance expression whereas the 3' MAR likely provides a dominant tempering activity. Interestingly a candidate fertility associated mutation in this locus has been mapped to the 3' MAR (Kramer et al. 1997). In other systems, association with the nuclear matrix is transient, selective and dynamic, paralleling the differentiative state and function (Heng et al. 2001; Heng et al. 2004; Ostermeier et al. 2003). Whether both are representative of a haploid expressed domain remains to be determined.

Conclusions

Spermatogenesis is an elaborate differentiative program that persists throughout the adult life of the male. The relative ease that cells from virtually every stage in the program can be isolated makes it an ideal developmental model. At the heart of this process is the repackaging of the sperm genome by the systematic replacement of histones with testes-specific variants - transition proteins - then their coordinated replacement by the protamines. The protamine cluster serves as a model for a haploid-expressed gene cluster whose elaborate regulatory program involves interactions with the nuclear matrix (Kramer & Krawetz 1996; Martins et al. 2004) as it is regulated by a suite of transcription factors (reviewed in Lu et al. 2006) and then post-transcriptional associations as RNPs (Fajardo et al. 1997). What remains to be determined is defining what forms of epigenetic modification are utilized to initiate and complete potentiation, i.e., the opening of a chromatin domain to enable and facilitate transcription. This underscores the pivotal role of chromatin remodeling (reviewed in Cho K. S. et al. 2004) that has emerged as a central theme towards understanding the mechanisms of disease (reviewed in Kleinjan & van Heyningen 2005). Perhaps, the disruption of these epigenetic marks at the paternally imprinted H19 locus are associated with oligozoospermia (Marques et al. 2004).

Although basal transcription can proceed from within structures of higher ordered chromatin (Georgel et al. 2003), there is a consensus that an open-potentiated chromatin structure is
required for optimum transcription. Potentiation, that is the opening of chromatin domains, grants trans-acting element access to transcription promoting cis-regulatory sequences. The potentiative timing in mouse (Kramer et al. 1998) and the steadfast nature of the human domain (Kramer et al. 2000) have been established. What is lacking are the crucial factor(s) that set potentiation in motion. A greater understanding of the interplay of nuclear constituents, including the nuclear matrix and the factors that can recruit many of these activating and silencing terms (Yasui et al. 2002) is required. This understanding will shed light on the regulation of this cluster and the processes the cell has at its disposal for the discrete selection of subregions of the genome for activity.

Acknowledgements

This work was supported by NICHD grant HD36512 to SAK.

References


Bridger JM, Boyle S, Kill IR & Bickmore WA 2000 Remodelling of nuclear architecture in quiescent and senescent human fibroblasts. Current Biology 10 149-152.


Repackaging the sperm genome 11


Ward WS, Kimura Y & Yanagimachi R 1999 An intact sperm nuclear matrix may be necessary for the mouse paternal genome to participate in embryonic development. *Biology of reproduction* 60 702-706.


