Imperial College London

Influence of sex chromosome complement and hormones on heterochromatin silencing in mice

Raquel Sofia André Santos Silva

Supervisor: Professor Richard Festenstein

Gene Control Mechanisms and Disease, Department of Medicine

A thesis submitted in partial fulfilment of the requirements for the degree of

MPhil in Clinical Medical Research

Imperial College of Science Technology and Medicine

September 2015

Declaration of Originality

The work presented was done written and executed by the author of this thesis, unless stated otherwise, where the contributions are acknowledged accordingly.

Copyright Declaration

Table of Contents

Table of figures

Table of tables

ABBREVIATIONS

bp	base pair
cDNA	complementary deoxyribonucleic acid
DAPI	4',6'- diamidino-2-phenylindole
DE	differentially expressed
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
EDTA	ethylenediaminetetraacetic acid
FBS	fetal bovine serum
FCG	four core genotype
GFP	green florescence protein
HDAC	histone deacetylase
HP-1	heterochromatin protein 1
K	Lysine
kb	kilo base
kd	Knockdown
KDM5	lysine demethylase 5
LINE	long interspersed nuclear element
M	Molar
me2	di-methylation
me3	tri-methylation
mRNA	messenger ribonucleic acid
PBMC	peripheral blood mononuclear cell
RNA	ribonucleic acid

SCC	sex chromosome complement
SEM	standard error of the mean
SINE	short interspersed nuclear element
PBS	phosphate buffered saline
qRT-PCR	quantitative real-time polymerase chain reaction
µg	micro gram
µl	micro litre
ºC	degree Celsius

ABSTRACT

Gender differences are accounted for by sex hormones, autosomal genes, sex-chromosome genes. Previously, the use of sex-reversed mice -hence producing XY females and XX males - revealed that a heterochromatic transgene and 369 autosomal genes in the thymus were differentially expressed dependent on the sex chromosome complement (SCC). These genes were enriched for sensitivity to HP1β dosage. Pericentromeric repeats were also less expressed in XY males when compared to XY gonadectomised/XX males and XY/XX females, suggesting a connection between testosterone, Sry and the Y chromosome.

Stemming from the above findings, it was sought to determine:

i. if the Y chromosome could enable testosterone to silence heterochromatic repeats
ii. whether removal of testosterone affected heterochromatin structure and integrity.

Since proximity to heterochromatin had been shown to influence expression of translocated genes and transgenes, the distance of the SCC genes to the nearest annotated major satellite sequence was determined. Twelve predicted protein encoding genes were found embedded in major satellites located near the centromere of chromosome 9. These genes were found to be repressed by testosterone on the XY male, which suggested a potential heterochromatin position effect operating on these endogenous genes which was regulated by sex hormones, tending to equalise their expression between the sexes.

Moreover, a trend of increasing number of heterochromatic foci per cell in circulating lymphocytes after gonadectomy of male mice was observed. Together with the upregulation of heterochromatic repeats in the thymus from gonadectomised male mice, suggests that absence of testosterone in males could affect genome stability. Lymphocytes derived from lymph nodes and thymus, showed tissue specificity and dependency on age in the expression of heterochromatic repeats and number of heterochromatic foci per cell. This data has implications for understanding how heterochromatin silencing is regulated between genders and has implications for understanding sex bias genome stability.

1-INTRODUCTION

1.1-Mammalian sex determination

Differences between genders are defined by phenotypic manifestations in anatomy, physiology and behaviour (Mittelstrass et al., 2011; McPhie-Lalmansingh et al., 2008; Ellegren & Parsch, 2007). These first appear at the very early stages of embryonic development and are heightened, later on, by gonadal differentiation (reviewed in (Arnold, 2012).

The prime triggers responsible for sex differentiation in therian mammals initiate before hormonal secretion, by expression of sex chromosome linked elements, such as the X-linked long non-coding RNA Xist which is necessary for female development (further discussed), and the Y-linked sex-determining region Y (Sry) gene, that acts as the master transcription factor in male gonadal development at 10.5-12.5 days postcoitum (dpc). Sry binds to the TESCO enhancer region of Sox9, initiating its transcription in Sertolli cells, which leads to the cascade for testis development and production of the Mullerian hormone that together eliminate the Mullerian duct. On the contrary in females, the Mullerian ducts develop into oviduct and uterus. Once functional, gonadal secretions further maintain and contribute to sexually dimorphic differences (reviewed in Silkaites & Lemos, 2014).

The main elements responsible for gender determination can be divided according to Arnold (Arnold, 2009) into: the (i) activation and (ii) organisational effects of hormones, and (iii), differential gene expression. Namely, activation effects are the ones that are eliminated in the absence of hormones, whereas organisational effects are the ones that remain stable even in sex hormone absence, such as morphological traits. On the other hand, differential gene expression can arise from (i) autosomal genes, including imprinted genes; (ii) X linked genes, in particular the genes that escape X inactivation in females; and (iii) specific elements from sex chromosomes, only expressed in one gender, such as Y linked genes and non-coding RNAs (ncRNAs) (further reviewed in (Arnold, 2014).

Burgoyne and collaborators also showed that pre-implanted male embryos had a higher developmental growth rate than female embryos of corresponding gestational time, suggestive of an early effect of Y linked genomic elements (Burgoyne et al., 1995). Moreover, transcription of thousands of genes was reported to be sex dependent in mouse and bovine blastocysts before gonadal differentiation (Bermejo-Alvarez *et al*., 2010; Kobayashi *et al*., 2006) which is consistent

with the idea that sex chromosomes, rather than phenotypic sex traits, play a key role in establishing gender differences.

Microarray analyses highlighted that interestingly thousands of autosomal genes are differentially expressed between genders also in adulthood, with a difference of ~1.2 fold for most of them (Yang et al., 2006). This is potentially of great importance in the light of the sex bias in the susceptibility towards various diseases, such as cancer, immunological, cardiovascular and brain disorders (Sarachana & Hu, 2013; Ji et al., 2010; Smith-Bouvier et al., 2008).

For example, Alzheimer's and autoimmune diseases such as systemic lupus erythematosis, rheumatoid arthritis and multiple sclerosis are more prevalent among women (Zhao et al., 2000); on the contrary, cardiovascular disease, Parkinson's disease, colorectal cancer and hepatocellular carcinoma are more common in men (Ober *et al*. 2008).

Furthermore, the physiological response to treatment is also gender dependent (Mittelstrass et al., 2011), which might be explained by metabolic differences between males and females. Thus, even though mortality, morbidity and disease progression were shown to exhibit a strong sex bias, their underlying molecular mechanisms remain largely unclear (reviewed in Silkaitis & Lemos, 2014).

Moreover, it is believed that less than half of the number of human sexual development disorders might be explained by alterations in well-known sex-determining genes (Munger & Capel, 2012). Therefore, epigenetic mechanisms, which are responsible for dynamic gene regulation without changes ins DNA sequence may be key players in the genesis of sexual development, and therefore in development disorders (Piferrer, 2013).

1.1.1-X chromosome inactivation in mammals

In zygotes with two X chromosomes, random inactivation of one of the X chromosomes happens early in embryonic development (figure 1). Initially the non-coding RNA XIST is expressed from both X copies, until the unbalanced expression from one of the X chromosomes copies, initiates the coatings in *cis* of the X chromosome that will be inactivated (Xi); the coating initiates recruitment of repressive protein complexes such as the Polycomb group (PcG). The chromosome structure alters to a more condensed state (section 1.2.1), known as the Barr body, which moves to the peripheral compartment in the nucleus.

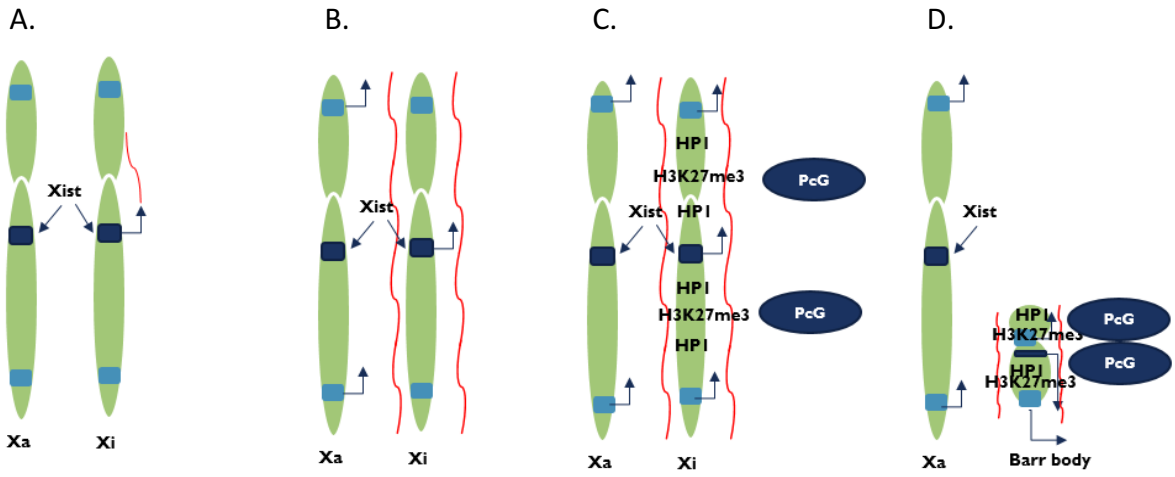


Figure 1 - Schematic representation of X chromosome inactivation in a mammalian XX background.

(A) random inactivation of one of the X chromosomes is initiated by higher expression level of the non-coding RNA XIST (B) which coats in cis the X chromosome that will be inactivated (Xi); (C) coating potentiates recruitment of repressive protein complexes such as the Polycomb group (PcG) that insert silencing marks; (D) Xi chromosome structure alters to a more condensed state - Barr body.

Unlike many sex biasing factors, X-chromosome inactivation serves to equalize gene expression in both genders. Indeed, any failure to silence one of the copies leads to embryonic lethality (Marahrens et al., 1997). However, a few genes escape from the inactivation process (further discussed). This is known to happen in 15% and 3% in human and mouse X chromosome genes, respectively (reviewed in Lemos, 2014).

Later defects in X-chromosome silencing after establishment in embryonic development are known to affect overall genome stability (Carone & Lawrence, 2012), including Barr body disruption that is a frequent feature in cancer cells (Barr & Moore, 1957); loss of Barr body structure is known to lead to overexpression of X-linked genes, and has also been identified in breast and ovarian cancers (Richardson et al., 2006; Jazaeri et al., 2002), suggesting these could potentially be involved in genesis of carcinogenic events.

1.1.2-Hormonal and gene expression interplay

Sex hormones are subdivided into: androgens, oestrogens and progesterone, all of which have cholesterol as a common precursor in their biosynthesis process. These hormones are mainly

produced in testis and ovaries, according to gender. Males have predominantly androgens, and oestrogens at low levels, whereas females have low levels of androgens and higher levels of the remaining classes. Steroid conversion also takes place in the adrenal glands, liver and fat tissue.

Hormones are powerful chemical messengers and key players in gene expression regulation (Ober *et al*., 2008). Their action is dependent on nuclear receptors specific to each class. Hormones reach their target cells through the blood stream and bind to their respective receptors that are located in the cytoplasm in an inactive form. The hormone receptors are then activated upon recognition of their substrate, which in turn initiates translocation of the hormone /receptor complex into the nucleus where it dimerises with another hormone/receptor pair. These dimers will then specifically bind to DNA sequences called hormone response elements (HRE), thereby acting as transcriptional activators/suppressors for downstream effectors of the hormone response (Sakiani, Olsen & Kovacs, 2012; Guerriero, 2009; Kosztin, Bishop & Schulten, 1997).

Many genes with metabolic and physiological functions have been identified as being controlled by steroid hormones in a gender-specific fashion. This includes fat distribution, liver metabolism, brain development, immune response and skeletal muscle development (Brown & Spencer, 2012; Sakiani, Olsen & Kovacs, 2012; Pasquali, 2006; Roy & Chatterjee, 1983). In particular, oestrogens have an immune-stimulatory activity, which is thought to be involved with auto-immune response, and to have a protective cardiovascular role (Herrmann et al., 2010; McMurray, 2001). This is in keeping with the higher susceptibility of females to immune diseases, and with higher cardiovascular prevalence disorders in males. On one note, women after menopause have the same cardiovascular risk as men (Jousilahti et al., 1999).

Interestingly, gene expression differences between genders are radically decreased after gonadectomy, reiterating how sex hormones are an important underlying cause for sexual dimorphism (van Nas et al., 2009).

1.1.3-Sex chromosome complement dosage

Sexual dimorphism entails a complex cross talk between sex hormones and genetic factors, such as the composition of the chromosome sex pair. The number and type of this pair is defined as the sex chromosome complement (SCC), which is usually made of XX in females, and XY in males in therian mammals. In females, even though one of the X chromosomes is inactivated,

some genes escape silencing and are expressed at higher level in females than XY individuals that only have one copy of the equivalent gene; interestingly, some of these escapees are epigenetic effectors. On the other hand, Y chromosome genes are only expressed in males, which also includes specific epigenetic effectors. In all, SCC contributes for the sex bias in genetic expression by several different mechanisms that remain to be further elucidated

In the last decade many phenotypes have been reported to be SCC linked rather than gender linked. These include adiposity (Chen et al., 2012), risk for autoimmune diseases (Sasidhar et al., 2012), DNA methylation (Zvetkova et al., 2005) and Position Effect Varigation (PEV) (Wijchers *et al*., 2010).

PEV was first described in the 1930s in the *Drosophila* genus (Muller, 1930) as the stochastic inactivation of the normally euchromatic *White* gene when this was translocated in proximity of pericentromeric heterochromatin. The same variegating phenotype can be recapitulated placing a reporter gene close to a heterochromatic domain, such as the centromere or the telomere. Festenstein et al. developed a mammalian model for this phenomenon, in which the gene subject to position dependent variegation encodes the human version of the CD2 molecule (hCD2), a constitutive T cell surface marker (Festenstein et al., 1996). Flow cytometry analysis allowed for the assessment of its expression levels by virtue of a fluorophore-conjugated antibody that is specie-specific as the presence of the CD2 molecule was shown to faithfully correlate with its mRNA levels.

This experimental system led to the surprising observation that in thymocytes the extent of silencing of the hCD2 gene was higher in XY male mice than XX females (Wijchers *et al*., 2010), suggesting that differences in gender or in the SCC can affect the expression of a gene subject to heterochromatic silencing. The Four Core Genotype (FCG) model permitted further investigation of this phenomenon. This murine model stems from a natural deletion of the Sry gene (Lovell-Badge & Robertson, 1990) that produces a female with an XY background; on the other hand, introduction of an SRY transgene (Mahadevaiah et al., 1998) resulted in XX males (figure 2).

Importantly, it has been shown that these mice display levels of sex hormones that correlate with their gender rather than with the sex chromosome complement (Gatewood et al., 2006; De Vries et al., 2002), hence making this a powerful approach to disentangle the effects of sex hormones from those of sex chromosome complement.

Figure 2 – Scheme for the FCG genotypes and respective gender.

In blue are the males XY-$^{+SRY}$ and XX^{+SRY}; in pink are the females XX and XY-. XY- mice had a natural deletion of *SRY*, and males in blue have a *SRY* ($^{+SRY}$) transgene on chromosome 3.

Previous work in the Festenstein lab using a hCD2/ FCG cross showed that the difference in transgene silencing correlated with the number of X chromosomes rather than with the gonadal sex (Wijchers *et al.*, 2010). This was verified in thymus from the F1 generation at 8w by Fluorescense-activated cell sorting (FACS) analysis. The microarray expression profile in the same tissue showed that approximately one thousand autosomal genes were sensitive to SCC within each gender pair, of which 369 genes were in common (Figure 3).

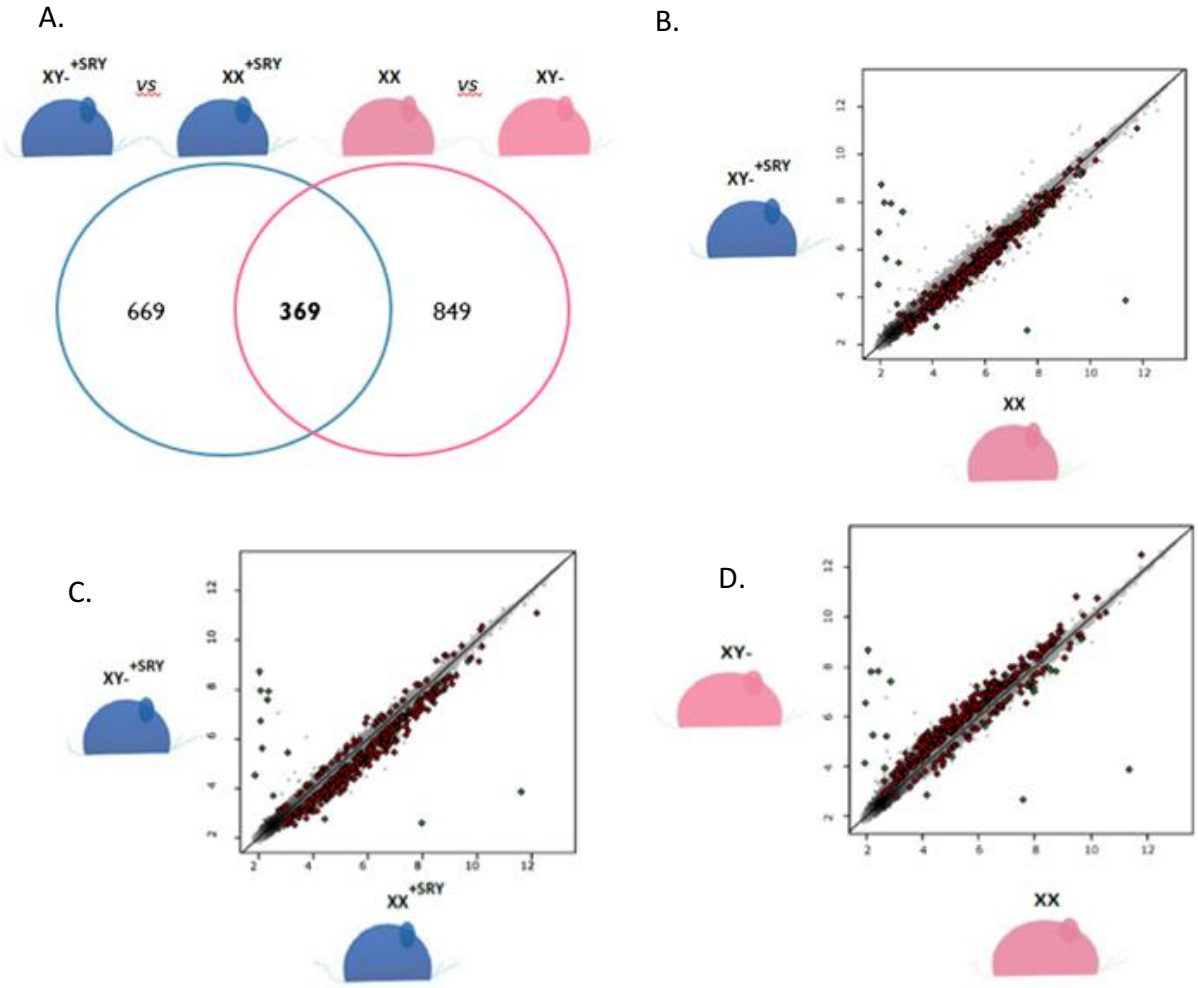


Figure 3 - Effect of SCC in autosomal gene expression, and modulatory role of SRY.

(A) Venn diagram representing differentially expressed genes between XX and XY males versus differentially expressed genes between XX and XY females (fold change>1.2, Student t test>0.05). The common 369 genes are therefore sensitive to the SCC, and are represented in the density plots in (B), (C) and (D), coloured in red; in green are the X and Y linked genes; the expression levels on each combination of genotype are plotted in a log2 scale where the diagonal represents no fold change (equal to 1). In (B) it is compared XX female with XY male, where XX expresses at lower level the majority of SCC sensitive genes. In (C) it is compared male XY versus male XX, where the SCC sensitive genes are mostly expressed higher in the XX background. In (D) it is compared female XX with female XY, where most of SCC sensitive genes are expressed higher in XY females. (C) and (D) suggest a regulatory role of Sry and/or gonodal hormones, directly or from its effects. Adapted from Wijchers *et al*., 2010.

The comparison of males XX^{+SRY} and XY^{-+SRY} showed this subset of sexually dimorphic 369 autosomal genes, were less expressed on an XX background compared to the XY background in the presence of SRY (Figure 3), supporting the idea that the presence of two X chromosomes has a silencing effect in this context. However, this profile is reversed when comparing XX and XY- females, suggesting that these differences are due to Sry and/or its indirect effects (Wijckers *et al*., 2010). Gene ontology analysis revealed enrichment for genes with metabolic function, suggesting an interesting link between epigenetic regulation and metabolism. Interestingly, there was enrichment within the subset of 369 SCC sensitive genes that were also sensitive to alterations in the levels of heterochromatin protein 1 β (HP1β), an epigenetic effector that promotes chromatin condensation (Zeng, Ball Jr & Yokomori, 2010).

Altogether, these results suggest that both hormones and sex chromosomes are crucial to fine-tune the expression of sexually dimorphic genes. Whether there is co-operation or redundancy between sex chromosome effects and hormonal effects remains a key unanswered question.

1.2- Chromatin structure and epigenetic modifications

Epigenetic mechanisms affect gene expression status without changes in the underlying DNA sequence. Within the nuclei of eukaryotic cells the DNA and its associated proteins form a highly ordered 3D structure termed chromatin. In multicellular organisms all cells have the same genomic material but not all genes are equally expressed. A finely-tuned network of transcription factors and epigenetic regulators are thought to lead to changes in potency and dictate cell differentiation and commitment. During the process of sex determination and gonadogenesis the pattern of gene expression is subject to a strict spatiotemporal regulation (Piferrer, 2013). In fact,

the bipotential embryonic gonads develop according to the action of antagonistic signalling pathways and transcriptional networks (section 1.1).

Importantly, epigenetic modifications such as histone modifications and DNA methylation can be passed on to daughter cells through mitotic divisions and are crucial for the acquisition and maintenance of cell identity as the developmental programme of the organism unfolds (Zaidi et al., 2010). However, how and which of these phenomena directly contribute to dimorphic differences between genders still remains poorly understood.

Histone modifications and chromatin folding

The fundamental unit of chromatin is the nucleosome, which is formed by 145-147bp of DNA wrapped around eight small basic proteins called histones (Luger et al., 1997a). The histone octamer is made up of two of each of the core histones H2A, H2B, H3 and H4 (Luger et al., 1997b). The histone H1 is located outside of the octamer and helps regulate chromatin compaction by binding between nucleosomes. The core histones -with exception of the histone H4- have evolved histone variants, which can vary from their canonical counterparts from only one or few amino acids (eg. H3) to a new domain inserted (eg. macro H2A). These variants play specific biological functions depending on the tissue or on the organismal developmental stage (Pusarla & Bhargava, 2005).

The N-termini of the histones protrudes outside of the nucleosome and can be targeted by the post-translational addition of a chemical group (Strahl & Allis, 2000). The most common histone modifications are methylations, acetylations, phosphorylations, ubiquitylations and SUMOylations. Their presence can affect DNA compaction either directly or by serving as a platform for the binding of other factors that recognise a given mark. An example of the first case is the negatively charged acetyl group, which reduces the affinity of histones to the also negatively charged DNA. This in turn leads to a more relaxed chromatin state that becomes more accessible for transcription factors and is believed to promote gene expression (Grunstein, 1997).

Notably, the same histone mark can have opposing effect according to the residue that is modified (Kouzarides, 2007). While some methylation events (e.g. H3K4me, H3K36me) facilitate gene expression, others like H3K9me, H3K27me or H4K20me determine the opposite outcome. Interestingly, it has been shown that mice that were deficient for the H3K9 demethylases also displayed reduced Sry expression and such XY embryos developed phenotypically as fertile

females (Kuroki et al., 2013). This report, while suggesting a direct regulation of the Sry locus by this class of chromatin remodellers suggested a critical contribution of histone modifications to the determination of sex in mammals.

Because histone modifications can occur alone or in combination with others on different residues, the existence of a "histone code" has been put forward (Jenuwein & Allis, 2001). This code proposed to be "read" by particular proteins via specific binding domains, and "written" and "erased" by other proteins that act on changing these modifications. In fact, these chromatin features were proposed to act as sensors for environmental and metabolic cues (Turner, 2009) through which a functional signal is orchestrated and translated into transient changes in gene expression or the maintenance of differential patterns in gene expression (Pedersen & Helin, 2010). Remarkably, histone modifications are reversible which creates a highly dynamic state in chromatin mediated gene regulation and expression (Pedersen & Helin, 2010; Shi et al., 2004).

DNA methylation

The DNA methylation process consists of the addition of a methyl group to a cytosine base, more common in 5`CpG3' dinucleotide sequences in mammals. This covalent process is catalysed by DNA methyltransferases (DNMTs). DNMT1 is responsible for maintenance of this epigenetic mark across cell divisions, meaning that it targets the hemimethylated bases on the new replicated strand, maintaining the heritability of this mark through cell division. DNTM3a/b target methylation on unmethylated CpGs – *de novo* methylation. The genome contains areas of high CpG content, named CpGs islands. These regions are usually located near promoters or regulatory elements where the change of methylation status is associated with gene expression regulation. Interestingly, DNMT3a and DNMT3b have been shown to be differentially expressed between male and females blastocysts which also correlated with differences in DNA methylation (Bermejo-Alvarez *et al*., 2008).

The somatic genome is highly methylated in a bimodal pattern. It is hypomethylated in CpGs Islands, transcription starting sites (TSS) and enhancer regions. On the other hand it is hypermethylated on non-CpGs-islands CpGs, gene bodies and repetitive regions (Cedar & Bergman, 2012). Its function is not fully understood but is has been observed that areas of DNA repetitive elements with loss of DNA methylation are particularly associated with structural alterations that could lead to chromosome destabilisation and consequently, cancer or other

genetic disorders (Narayan et al., 1998). This loss is also frequent in ageing processes (Gentilini et al., 2013).

RNA interference

The RNAi pathway acts at the post-transcriptional level on mRNA transcripts by short RNA sequences of 24 bps that are able to silence a particular gene by promoting cleavage of its transcript in coordination with specific protein complexes such as Dicer (Elbashir et al., 2001).

This highly conserved mechanism has been described for fission yeast, plants and more recently to a much more limited extent in mammalian cells (Volpe & Martienssen, 2011). In *S.Pombe* this mechanism serves as trigger for heterochromatin formation, where the heterochromatic transcription initiates production of these short RNAs and recruits enzymes involved in histone and DNA alterations (Noma et al., 2004).

Short dsRNAs have been further reported to target CpGs islands in promoter regions, that are able to induce transcriptional silencing by histone and/or DNA methylation in mammalian cells (Castanotto et al., 2005; Morris et al., 2004).

It has been suggested that DNA methylation and RNAi could represent primitive forms of a defensive mechanism against DNA and RNA viruses to allow the recognition and subsequent neutralization and silencing of the intruders. However, due to the evolution of the chromatin structure and the appearance of histone modifications, the two systems subsequently diverged and gained their own peculiar features (Matzke, Mette & Matzke, 2000).

1.2.1- Euchromatin and Heterochromatin

In the nuclei of eukaryotic cells two distinct forms of chromatin can be observed: Euchromatin and Heterochromatin. While euchromatin represents the relaxed state of chromatin and harbours the vast majority of actively expressed genes, heterochromatin is more tightly packed and transcriptionally inert (Dillon, 2004).

A number of protein complexes contribute to the determination of the chromatin state in response to metabolic and environmental cues. This is a dynamic process in which the factors promoting chromatin relaxation are antagonised by those with the opposite function (Figure 4).

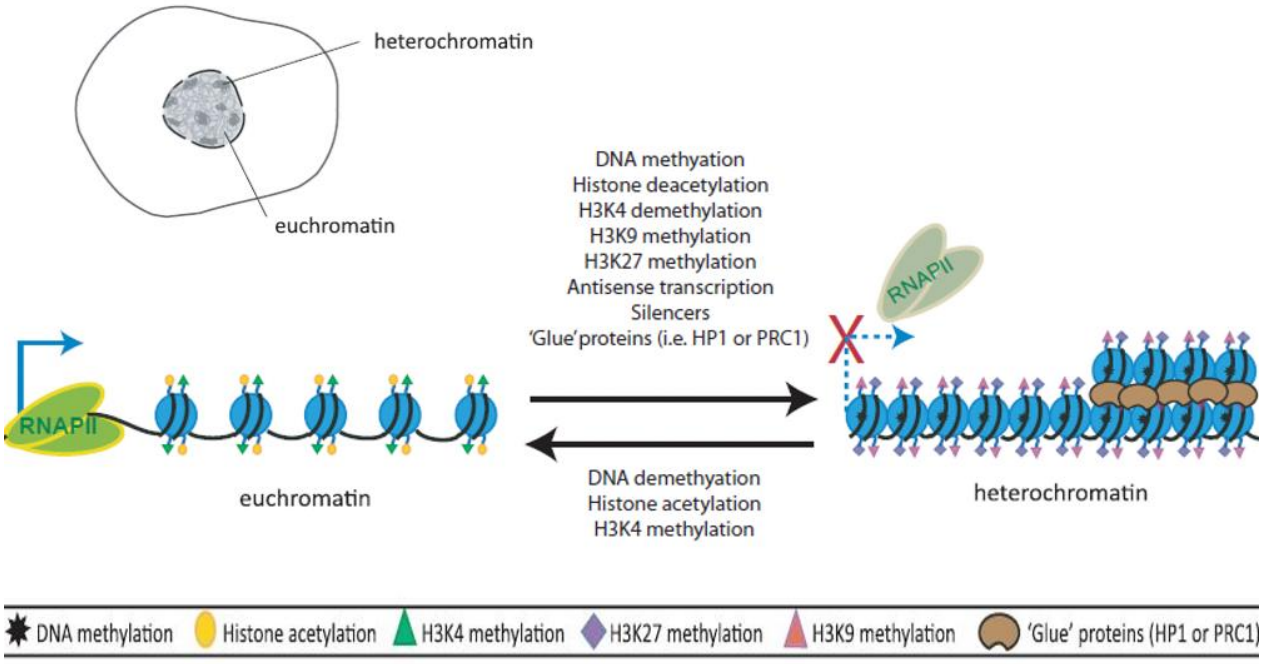


Figure 4 - Switch triggers leading to transition between euchromatin and heterochromatin.

Euchromatin is associated with histone acetylation and H3K4me, whereas, DNA methylation, histone deacetylation, H3K4 demethylation, H3K9 and H3K27 methylation are typical markers of heterochromatin. Heterochromatin Protein 1 (HP1) and Polycomb repressive complex 1 (PRC1) also contribute to the determination of the chromatin state Adapted from (Yandim, Natisvili & Festenstein, 2013).

1.2.2- Heterochromatin forms

Heterochromatin can be further subdivided into constitutive and facultative: (i) domains of constitutive heterochromatin regions, are associated with tandem repetitive DNA sequences and are frequently located in pericentromeric and telomeric regions (figure 4); on the other hand (ii), facultative heterochromatin forms in response to developmental cues, giving rise for example, to the Barr body during the X chromosome inactivation in female mammals (Grewal & Jia, 2007).

A significantly large proportion of the mammalian genome is occupied by heterochromatic repetitive DNA repetitive elements (70% in human and 50% in mouse genome). Ensuring transcriptional silencing of DNA repeats by heterochromatinisation is essential for maintaining genome and cell integrity, as transcriptional derepression of such regions has been associated with aging and development of diseases such as cancer (Schulz et al., 2009; Barbot et al., 2002). Nonetheless, recent reports have shown evidence of a low percentage of active centromeric

transcripts, thought to play a role in heterochromatin formation, cell division, differentiation and genome stability (Hall, Mitchell & O'Neill, 2012).

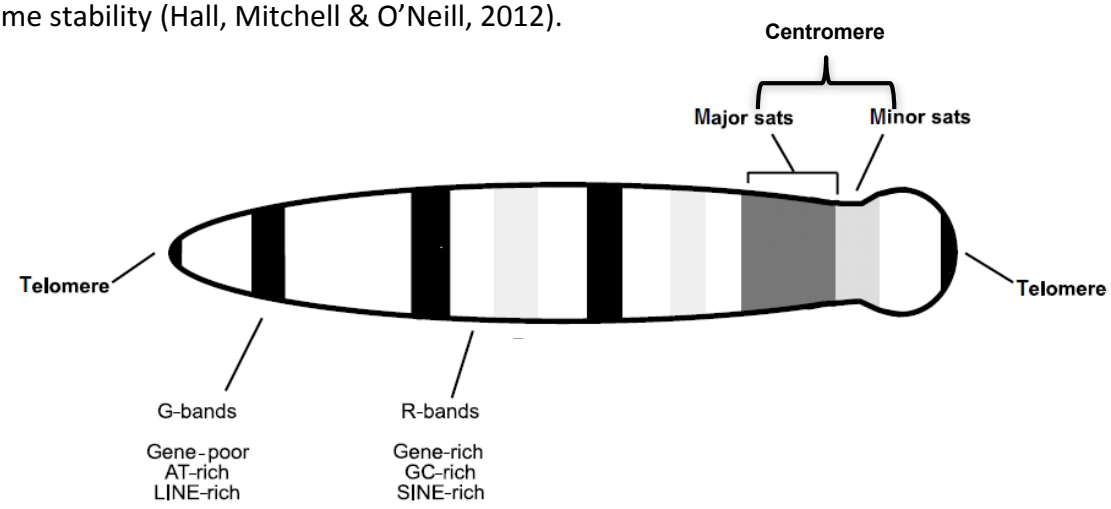


Figure 5 - Schematic representation of the distribution of tandem repetitive elements in a mitotic mouse acrocentric chromosome.

Constitutive heterochromatin is located in centromeres and telomeres. In centromeric and pericentromeric regions, are located the minor and major satellite tandem repeats, respectively. Other classes of repetitive elements include interspersed DNA transposons and the Retrotransposons, LINE and SINE, which are located in G and R bands respectively. Adapted from (Martens et al., 2005).

Mouse pericentromeres consist of AT-rich regions, with repetitive sequences that are 234 bp long (aka, gamma-satellite or major satellite repeats) (Martens et al., 2005). During interphase, these regions on different chromosomes associate and form the so-called heterochromatic clusters (or "chromocenters"), which can be visualised as distinctive foci upon 4',6-diamidino-2-phenylindole (DAPI) staining, which is a fluorescent stain that binds strongly to A-T rich regions in DNA. These regions are enriched for heterochromatic marks, namely histone modifications like H3K9me3, H4k20me3, H3K27me1, H3K64me3, as well as DNA methylation and have reduced sensitivity to nuclease treatments (Lange et al., 2013).

The pericentromeric tandem repeats serve as a barrier between euchromatin and centromere cores, therefore representing an essential element for the maintenance of centromere integrity and function (Chen et al., 2008). They are also important to avoid recombination events in these regions during cell division (Ando et al., 2002). Centromere structure is fundamental for kinetochore assembly and subsequent chromosome segregation. Constitutive heterochromatin on DNA satellite repeats provides an excellent model to uncover global mechanisms of heterochromatin structure and function (Bannister & Kouzarides, 2011; Dillon, 2004). Recent studies in murine models reported major and minor satellite transcripts, located in

pericentromeres and centromeres, respectively, that are thought to be involved in heterochromatin maintenance (Hsieh et al., 2011). Satellite derepression and upregulation has on the other hand been identified under stress conditions, aging and oncogenesis (Hsieh et al., 2011; Zhu et al., 2011; Frescas et al., 2008; Jolly et al., 2004) .

1.2.2.1- The role of HP1 in heterochromatin formation

As mentioned, chromatin condensation and silencing entails a stepwise process which requires nucleosomes to be decorated by specific histone modifications. To form pericentromeric regions in mice, H3K9me1 is converted into H3K9me3 by the methyltransferases Suv39h1 and Suv29h2. The exact mechanism is still unknown but it is clear that the H3K9me3 enrichment is responsible for further recruitment of specific proteins involved in chromatin binding, such as the modulator heterochromatin protein 1 (HP1). Mammals have three isoforms, namely, HP1α, HP1β and HP1γ, of which the first two are frequently associated with pericentromeres, exhibiting relatively redundant functions (Maison et al., 2002).

HP1 is known to potentiate heterochromatin by recruiting other proteins with catalytic functions to chromatin and also facilitates H3K9me3 spreading due to its association with Suv29h (Yamamoto & Sonoda, 2003) in a positive feedback loop. The “closed” structure is further stabilized by recruitment of the methyl-CpG-binding protein MECP2 and its association with either deacetylase or methyltransferase activity, both associated with establishing silencing marks (Lange et al., 2013).

HP1 has an essential role in cell integrity, as knockdown of this protein results in abnormal heterochromatin structure and loss of its functional activity (Kwon & Workman, 2011) and it has been shown that pericentromeric chromosome clusters colocalised with the Green Fluorescent Protein signal of HP1β-GFP proteins (Festenstein . 2003).

1.2.3- The mammal repeatome

Complete sequencing of the genomes of hundreds of organisms revealed that their degree of complexity showed no correlation with the number of protein-coding genes (Taft, Pheasant & Mattick, 2007). On the contrary, non-coding sequences occupy large proportions of the genomes

of more complex organisms, and it was somehow surprising that these sequences are species-specific.

Table 1 – Classification and features of repetitive elements in the human genome.

(Adapted from (Casa & Gabellini, 2012).

<table>
<tr><th colspan="4">Repeat type</th><th>Average repeat length</th><th colspan="2">Genome coverage</th></tr>
<tr><td rowspan="4">Interspersed</td><td rowspan="3">Retrotransposons</td><td colspan="2">Long Terminal Repeat (LTR)</td><td>6-11 kb</td><td>8%</td><td rowspan="3">42%</td></tr>
<tr><td rowspan="2">Non-LTR</td><td>LINE</td><td>6 kb</td><td>20%</td></tr>
<tr><td>SINE</td><td>0.3 kb</td><td>13%</td></tr>
<tr><td colspan="3">DNA transposons</td><td>1-3 kb</td><td colspan="2">2-3%</td></tr>
<tr><td rowspan="6">Tandem</td><td rowspan="3">Satellite</td><td colspan="2">Alpha (unit length: 171 bp)</td><td>3-5 Mb</td><td colspan="2" rowspan="6">22-25%</td></tr>
<tr><td colspan="2">Beta (unit length: 68 bp)</td><td>2-14.5 kb</td></tr>
<tr><td colspan="2">Gamma (unit length: 220 bp)</td><td>10-200 kb</td></tr>
<tr><td rowspan="3">VNTR (Variable number of tandem repeats)</td><td colspan="2">Microsatellite (unit length: 1-13 bp)</td><td>Hundreds bp</td></tr>
<tr><td colspan="2">Minisatellite (unit length: 6-100 bp)</td><td>1-15 kb</td></tr>
<tr><td colspan="2">Macrosatellite (unit length: 2-13 kb)</td><td>Up to hundreds bp</td></tr>
<tr><td colspan="4"></td><td><u>Total</u></td><td colspan="2"><u>66-70%</u></td></tr>
</table>

The repetitive elements class accounts for up to 70% of the human genome and 44% of the mouse genome, (de Koning et al., 2011; Neguembor & Gabellini, 2010). Despite described

functions in chromosome structure, gene regulation, genome plasticity and evolution, repetitive elements genesis and mechanisms are poorly understood (Casa & Gabellini, 2012).

There are two main classes of repetitive elements, the interspersed and the tandem repetitive elements (table 1). The peculiar feature of the former category is their capacity to move within the somatic genome. These mobile elements present differential expression among tissues of the soma, and constitute a major source of genetic variations within mammals (Ekram et al., 2012; Muotri et al., 2005).

In order to maintain genome stability, expression of repeats has to be finely tuned to protect cells from aberrant activation of repetitive elements (Casa & Gabellini, 2012).

Indeed, more than 100 transposition events have been identified as causing diseases, including human cancers (Solyom & Kazazian Jr, 2012; Lee et al., 2012). Tandem repeat instability has also been shown to be the underlying cause of numerous malignant disorders (Lopez-Contreras & Fernandez-Capetillo, 2012).

1.3- T cell development

In mammals, all adults hematopoietic cell originate in the bone marrow. Here they acquire commitment to each cell lineage including lymphocytes. The ones that migrate to the thymus for maturation and selection are named T cells (Rothenberg and Taghon, 2005).

At the initial stage, the T cell precursors are named double negative (DN) as they do not express on their cell surfaces neither CD4 nor CD8 co-receptors. After the first selection, termed β-selection, only lymphocytes who have a functional T cell receptor β (TCR β) are allowed to proliferate and expand. Selected DN cells progress to the double positive (DP) stage, where both CD4 and CD8 receptors are expressed. At the same time, following rearrangements at the α and β chain loci, T cells express TCRs that have the capacity of recognising a wide range of antigens. DP thymocytes undergo new selection steps by exposure to the major histocompatibility (MHC) protein that carries a “self” antigen fragment. The cells that interact with the MHC complex are positively selected, and receive a survival signal, whereas the remaining and vast majority die by neglect (Kisielow et al., 1988).

In contrast, negative selection results in the elimination of self-reactive cells. In this event an antigen presenting cell (APC) that bears the MHC complex presents its antigen to T cells. The ones that recognise and strongly react to the self-protein are targeted for apoptosis (negative selection). Selected thymocytes undergo T lineage differentiation into single positives (SP) T cells, expressing either CD4 or CD8, hence becoming T helper (CD4+), and on the second, cytotoxic (CD8+), respectively. Afterwards, SP thymocytes are fit to migrate to lymph nodes through the blood stream where they will constitute the peripheral T cells pool.

Upon stimulation by dendritic cells (a specialised class of APC) during an immune response, these T cells can develop into mature effector T lymphocytes and aid in the response. Some of these cells further develop into T memory cells and are stored in the spleen, one of the peripheral lymphoid organs.

1.4-Hypotheses

(i) age, hormones and sex chromosomes modulate heterochromatin silencing;

(ii) heterochromatic repeats influence expression of nearby genes

(iii) testosterone helps maintain genome stability in males.

1.5-Aims

(i) identify fundamental mechanisms whereby heterochromatin regulates genes expression in mammals;

(ii) identify if the Y chromosome and testosterone co-operate to silence heterochromatin;

(iii) investigate how gonadectomy and SCC might affect heterochromatin structure and genome stability.

2-MATERIAL & METHODS

2.1- Cell Culture

Suspension cells were grown at 37°C in 5% CO_2 in RPMI with high glucose and without L-Glutamine (GIBCO ®) and no phenol red, 10% v/v knock Out serum (KSR) (Sigma), 1% v/v Antibiotic/Antimycotic (GIBCO ®) and 1% GlutaMAX™ (GIBCO ®). Culture medium was replaced every 3 days.

2.1.1- T-cell activation

Lymph nodes from C57BL/6 male mice were collected at 6 weeks old. Tissue was teased in cold PBS and strained (40µm). Cells were incubated for 30 min at 4ºC and rotating together with dynabeads Mouse panB for B cell depletion (50µl of beads/5x10^6 cells). Cells were separated from magnetic beads and were prepared for culturing. It was added T-cell activation beads (25µl of beads/1x10^6 cells) and human IL2 (3µl/ml). 1 ml of medium was added per 1x10^6 cells, and 0.250ml to each well in a 96-well plate.

2.2- Experimental Mice

C57BL/6 mice were acquired from Jackson Laboratories or Charles River and bred in the MRC Animal House in accordance with the UK Home Office law. The Festenstein lab had previously generated a transgenic mice strain, GM31.3 in which T cells express a GFP-HP1β chimeric molecule. The strain was obtained by fusing a cDNA construct under the control of a human CD2 locus control region (Festenstein et al., 2003).

For the gonadectomy/sham procedures, GM31.3 male mice were used, with 8-10 weeks. Gonadectomy were performed by the vet, according to established guidelines; sham consisted of placing the animals under anaesthesia and cut/stiches. Blood samples, thymus and lymph nodes were collected for T cell analysis using the ImagestreamX.

2.3- Total RNA extraction

Trizol (Invitrogen, USA cat. 15596-018) was used to extract RNA from mice thymus or from harvested cells. 0.5 to 1 ml of Trizol (Invitrogen) was used to re-suspend the cell pellet and incubated for 5 minutes at room temperature. Next, 200 µl of chloroform (Sigma) per 1 ml Trizol

was added, mixed by 20x inversions and incubated for 5 minutes at room temperature. Samples were then centrifuged at 12 000g for 15 minutes at 4°C and the upper aqueous phase (~ 500 µl) was transferred into new tubes. 500 µl isopropanol was added, mixed by vortexing and incubated at room temperature for 10 minutes; when the starting amount of cells was low, 50µg of glycogen (Invitrogen) was added to aid the RNA precipitation. The samples were centrifuged at 12 000 g for 15 minutes at 4°C. The supernatant was then removed and the RNA pellet was washed with 1ml of cold 75% ethanol. Samples were mixed by gentle vortexing/flick and centrifuged for 10 minutes at 8000 g at 4°C. Finally the supernatant was discarded and the RNA pellet was briefly air-dried for following treatment with DNase I with the DNA-free Kit (Ambion), according to protocol instructions. The supernatant was transferred into new tube and placed on ice. RNA concentration was assessed using NANODROP ® Spectrophotometer ND-1000 with ND-1000 v3.3.1 software. When necessary, samples were diluted to a concentration of 250ng/µl with DEPC treated water. The quality of the extracted RNA was also checked by running a 1% agarose gel.

2.4- cDNA synthesis

Using the ThermoScript™ kit (Invitrogen), RNA samples were reverse transcribed to cDNA. For each reaction 4.5 µg of RNA were mixed with 2 µl of random hexamers, 4 µl of 10 mM dNTPs; DEPC water was then added to make up a total volume of 26 µl. The reaction mixture was incubated 5 minutes at 65°C and then placed on ice. Subsequently 8 µl of cDNA synthesis buffer, 2 µl of 0.1 M dithiothreitol (DTT), 2 µl of RNAseOUT™ and 2 µl of ThermoScript™ reverse transcriptase (15 U/ µl) were added to each reaction mix. Samples were then incubated at 25ºC for 10 minutes, 50°C for 50 minutes, then 5 minutes at 85°C and finally 30 seconds at 4°C. 35 µl out of a total volume of 40 µl of the cDNA containing solution were transferred into new tubes containing 215 µl of DEPC treated water to make up a total volume of 250 µl.

2.5- Quantitative real time PCR analysis

Quantitative RT-PCR estimated the level of expression of specific mRNAs, using the SYBR® Green JumpStart™ Taq Ready Mix (Sigma cat. S4438). Each reaction had the composition of: 10 µl of SYBR® Green JumpStart™ Taq ReadyMix, 2.5 µl of 1 µM forward primer, 2.5 µl of 1 µM reverse primer and 5 µl of cDNA template. Low profile 96 wells white PCR plates (BIORAD) and transparent caps (MicroRad) were also used. The prepared plate was briefly spanned down (3 minutes, 2500

rpm at 4°C) before being loaded in the PCR machine. The PCR reaction was run on a Chromo4 DNA engine (MRJ) along with the Opticon Monitor 3 software (Biorad). The adopted programme for cDNA samples was 95°C for 2 minutes, 95°C for 30 seconds, 58°C for 30 seconds, 72°C for seconds, plate read, 80°C for 1 minute, plate read, 82°C for 1 second, plate read, 85°C for 1 second, plate read, go to step 2 for 40 cycles, following a melting curve from 70°C to 95°C, read every 0.5°, holding for 1 second at each temperature. At the end of the amplification reaction the amplification curves of all samples analysed were visualized on the Opticon Monitor 3 software. The cycle threshold (Ct) values were imported to the software to Microsoft Office Excel 2010, after checking the melting curves to estimate the specificity of the amplification. The calculations were performed using the average C(t) value obtained from 3 technical replicas, and were normalized against the amount of cDNA amplified from a housekeeping gene.

2.6- Chromosome cluster analysis from mice blood and lymph nodes samples

The ImageStreamX (Amnis, Seattle, Washington) is a combination of a flow cytometer and a digital fluorescence microscope that allows cells to be directly imaged individually in suspension. The procedure was used to image lymphocytes from collected blood, thymus and lymph nodes samples from gonadectomy/sham GM 31.3 mice. Blood samples were diluted in heparin and washed in 3 ml of cold PBS supplemented with 2mM EDTA and 0.05% of Sodium Azide; thymus and lymph nodes were collected and teased in the same supplemented cold PBS and cells were counted to obtain 0.5 million cells. Cells were centrifuged at 1500 rpm for 5 minutes at 4ºC. Supernant was sucked and 6-12 ml of pre-warmed Gey's Solution was added to lyse red blood cells. Samples were incubated at 37ºC until solution turned clear. Centrifugation at the same conditions followed. 100µl of 1:7500 diluted Draq5 (5µM) (in cold PBS) was added to the cell pellet. Samples were placed on ice and kept in the dark until imaging. 10 000 – 20 000 events were collected per experiment with EDF1 filter and 60x magnification. The laser wavelength for GFP signal was 488 nm, (intensity of 100.0) and Draq5 was 658 nm (intensity of 20.0). Results were quantitated virtually using the associated Image Data Exploration and Analysis software (IDEAS Amnis). After single cell and GFP/Draq5 fluorescence gating, the number of chromosome clusters in T-cells nucleus was determined by computing the intensity of localized bright spots within the image that were 2.75 pixels in radius. Since GFP protein is specific for T-cells, by gating out/in GFP

positive/negative cells, it was possible to look differentiate T and B cells. Results were transferred to Microsoft Office Excel 2013, and plotted for analysis.

2.7- Bioinformatic analysis of intergenic major satellite repeats

To assess the expression of repeats located near genes, a script was developed by Dr Thomas Carrol (MRC-CSC) in R to locate the coordinates of 42 major repeats (retrieved from RepeatMask) to nearby genes that were differentially expressed in the different genotypes of FCG, also comparing intact versus gonadectomised mice. Further visualization tools were used to look in these regions such as IGV and NCBI. Finally, because some of these repeat-proximal genes were protein encoding, NCBI tools helped assess potential protein domain structures coming from these genes.

3.1- Introduction

The repetitive nature of satellite elements makes their alignment to the genome difficult. An approach developed by Dr Felix Krueger (Babraham Institute) was applied to the data presented in this section that, in brief, aligns the sequencing reads to the tandem repeat monomer. Since microarray probes do not include repetitive regions, this methodology limits the analysis to the RNA-seq data.

Previous unpublished results from the Festenstein lab showed that heterochromatic repetitive satellite transcripts from RNA-seq data, were differentially expressed according to gender and to the presence of the Y chromosome and Sry in the FCG mice (M. Mauri, Master Thesis 2012). This result showed that in thymocytes of 10 months old XY males, both major and minor satellite repeats were strongly repressed when compared to XX males and to XY/XX females (Figure 6).

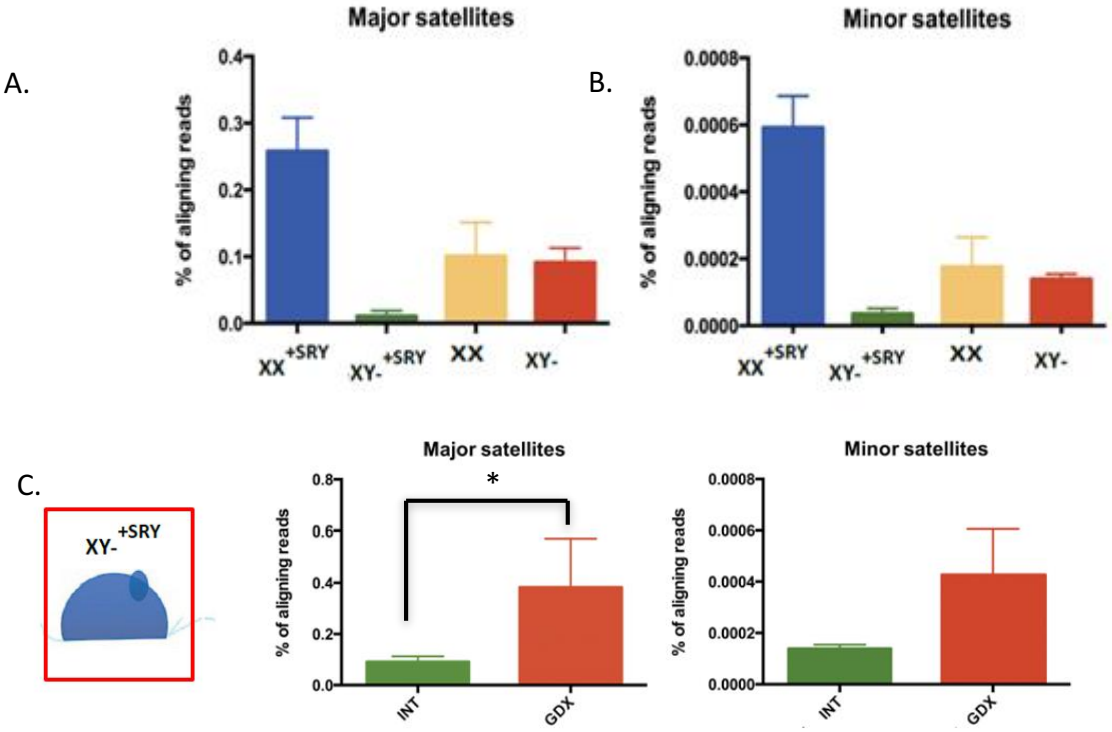


Figure 6 – Expression profile of major and minor satellites on the four different genotypes characteristic of the FCG model assessed via RNASeq.
Major satellite (A) expression follows the same trend as minor satellite (B). Their expression is consistently repressed in XY males, compared to XX and XY females, suggesting a gender dependency. Addition of Sry transgene in a XX background (male phenotype) leads to upregulation of these repeats suggesting a sex chromosome complement effect (upper panel). Upon removal of the gonads in XY males, the heterochromatic repeats were upregulated consistent with testosterone being directly implicated in heterochromatic silencing mediated by Sry and the y

chromosome (C). Statistical significance of the differential expression was assessed via an ANOVA one-way test. Differences observed in male XY background are significant (ANOVA test, p-value<0.05$^{(*)}$) in intact versus gonadectomised for major and minor satellite expression. Festenstein lab, M. Mauri, unpublished data).

Since XY males and XX males were shown to have similar testosterone levels (Gatewood et al., 2006; De Vries et al., 2002), it was suggested that in males the Y chromosome and the male hormones could cooperate in the transcriptional repression of DNA repetitive elements. To assess this hypothesis, the expression levels of satellite repeats were analysed in the thymocytes from 10 month old mice in XY males that had not been gonadectomised (intact) and XY males that were gonadectomised at 10 weeks of age. RNA-seq data showed that gonadectomised XY males upregulated the major and minor satellite repeats, suggesting testosterone, in combination with the Y chromosome and Sry, provided a potential mechanistic link for heterochromatin silencing.

In addition to repeats, a number of genes were also differentially expressed (DE) between the FCG genotypes. Significant fold changes in expressed genes were assessed with DeSeq (Bioconductor) (Anders, 2010) and those with p values less than 0.05 were considered differentially expressed.

Gene ontology analysis (GOSeq software) revealed enrichment for genes involved in the functioning of the heart and muscles, cell receptors and transporters, as well as those involved in apoptosis and lipid metabolism. Even though these GO terms were similar to what had been previously described in Wijchers *et al*. (2010) there was little overlap between the actual genes identified in these two studies. This discrepancy might be explained by the differences in the age of the mice, in the Wijchers study mice where 8 weeks old whereas in the more recent transcriptomic study they were 42 weeks old. Another explanation might be that they have different genetic backgrounds C57/Black6 (Jackson lab) in the recent study and F1 cross between C3H X C57Black6 with X chromosome fixed as C57Black 6 in the Wijchers study. Moreover the two studies used different techniques for assessing gene expression, RNA-seq or microarrays, that could also have contributed to the difference.

3.2- Role of major satellites in regulating autosomal gene expression

In the light of the observation that gonadectomised mice have higher thymic expression of repeats and also show dysregulation of a number of autosomal genes, it was hypothesised that

the two phenomena could be somehow coupled. In particular it was reasoned that the transcriptional de-repression of repetitive elements could in turn affect the expression of nearby genes. With this in mind, using the RNA-seq data set presented in the previous section, it was sought to calculate the distance between any differentially expressed gene and the nearest major satellite.

By using the RepeatMasker tool on the UCSC (University of California Santa Cruz) browser, the coordinates of 42 major satellites rich intergenic regions that mapped to the mouse genome (*Mus Musculus* genome assembly version 9 – mm9 file) were retrieved (Table 2).

Table 2 – Major satellite repeat sequences in mm9, annotated using RepeatMask.
In red is marked the location of a major satellite block which co-localises with one of the differentially expressed genes the analysis of which is present below (figure 10).

Chr – Coordinates – Satellite- length-strand
chr2 98502391 98502657 GSAT_MM 2001 -
chr2 98502984 98503150 GSAT_MM 1201 +
chr2 98503798 98504248 GSAT_MM 2277 -
chr2 98506701 98507489 GSAT_MM 3461 +
chr3 56379356 56379673 GSAT_MM 1166 -
chr3 99785017 99785317 GSAT_MM 877 -
chr4 3016701 3018331 GSAT_MM 1502 +
chr4 70039150 70039288 GSAT_MM 779 +
chr5 104585506 104585557 GSAT_MM 293 +
chr5 115372463 115372611 GSAT_MM 454 -
chr6 43293760 43293805 GSAT_MM 281 +
chr6 60637216 60637324 GSAT_MM 479 +
chr8 24205947 24206277 GSAT_MM 1271 +
chr9 3000002 3038419 GSAT_MM 3299 -
chr9 35112790 35113194 GSAT_MM 2737 +
chrX 72889685 72891547 GSAT_MM 1283 -
chrX 72892257 72893405 GSAT_MM 1282 -
chrX 72893356 72897442 GSAT_MM 1374 -
chrX 72897455 72898719 GSAT_MM 1223 -
chrX 72899429 72900279 GSAT_MM 1166 -
chrX 72900164 72902454 GSAT_MM 1222 -
chrX 72902391 72903323 GSAT_MM 1172 +
chrX 72904033 72905299 GSAT_MM 1223 +
chrX 72905312 72909396 GSAT_MM 1374 +
chrX 72909472 72912683 GSAT_MM 1357 +
chrX 72912747 72913319 GSAT_MM 1262 +
chrX 72913327 72914566 GSAT_MM 1276 +
chrX 72914502 72915573 GSAT_MM 771 +
chrX 72915665 72916807 GSAT_MM 1002 +
chrX 72916791 72917881 GSAT_MM 939 +
chr11 3000013 3002238 GSAT_MM 1614 -
chr11 3002257 3004752 GSAT_MM 1529 -
chr12 3109871 3110121 GSAT_MM 1924 +
chr12 48934921 48935109 GSAT_MM 663 +
chr12 68160158 68160943 GSAT_MM 1844 +
chr12 75904824 75905492 GSAT_MM 1691 -
chr13 6403879 6403989 GSAT_MM 355 +
chr13 77578170 77578318 GSAT_MM 1087 -
chr16 10975160 10975376 GSAT_MM 816 -
chr18 3007516 3007667 GSAT_MM 506 +
chr18 56099124 56099300 GSAT_MM 771 -
chr19 23431912 23432083 GSAT_MM 825 +

The distances between the nearest intergenic major satellite repeats and differentially expressed genes were calculated and binned in different groups according to distance (Figure 7, 8). The different genotypes comparisons identify specific genes groups, which are highlighted in table 3.

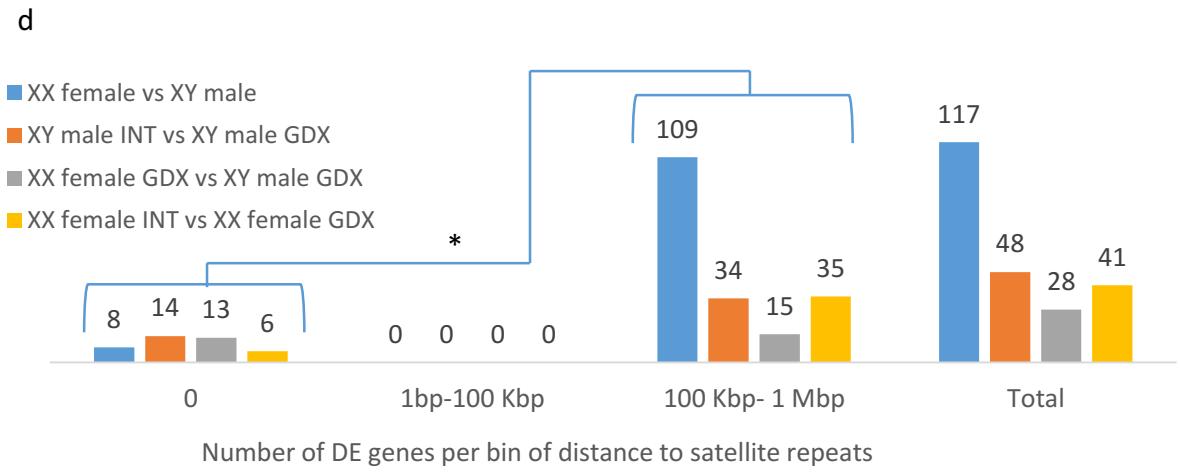
The group with highest number of DE genes in total was (i) XX female vs XY male, which showed a total of 117 DE genes between males and females, from which 8 (7%) had a distance of 0bp to the repeats.

Table 3 – Genotype comparisons highlight different DE gene groups

Genotype comparison	Group of DE genes
XX female INT vs XY male INT	Sex dimorphic genes
XX female GDX vs XY male GDX	SCC sensitive genes (in the absence of hormones)
XX female INT vs XX female GDX	Female hormone sensitive genes
XY male INT vs XY male GDX	Testosterone sensitive genes

Upon gonadectomy (ii), the number of DE genes was decreased significantly to 28 DE genes, which as expected, suggests that hormones increase differences in gene expression between genders. However, on the other hand, the number of DE genes that were proximal to major satellites was increased from 8 to 13 DE genes upon gonadectomy of both males and females (46% of total DE genes). This suggests that sexual hormones reduce the difference in expression between male and female of these major satellite proximal DE genes as gonadectomy enhances the male vs female difference. Within-gender comparisons of (iii) intact XY males vs gonadectomised XY males, showed a total of 48 DE expressed genes of which 14 (29%) had 0bp distance to major satellite repeats; on the other hand, (iv) intact XX females vs gonadectomised XX females showed a total of 41 DE genes of which 6 (15%) were proximal to major satellites.

Very interestingly, the 6 major satellite proximal DE genes from the comparisons (iv) is a subset of the 8 major satellite proximal DE genes from (i), which are included in the subset of major satellite proximal 13 DE genes from (iii), that finally are included in the subset of major satellite proximal 14 DE genes from (ii), and they were all testosterone sensitive. This suggests that testosterone regulates expression of these major satellite repeats, which in turn controls the genes nearby (see discussion 4.1).



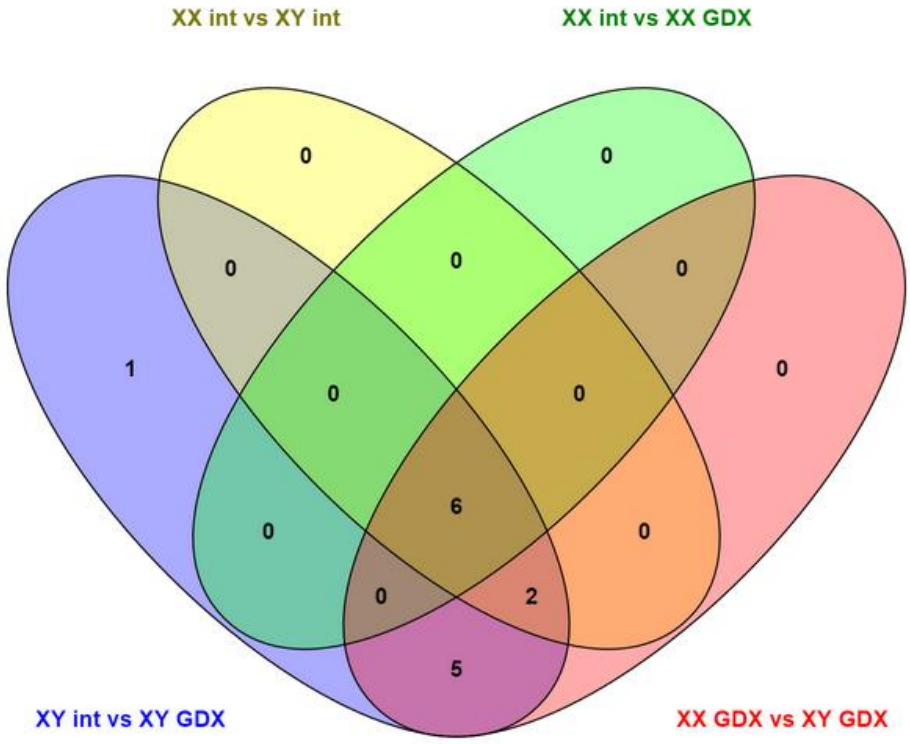


Figure 7- The subset of differentially expressed genes in XY males and XX females, intact (INT) versus gonadectomised (GDX) in the FCG are located subjacent to major satellite repeats.

Top: The distribution of differentially expressed genes according to distance of nearest major satellite repeat revealed a higher proportion of hormone sensitive genes at 0 bp distance to major repeats, compared to sexually dimorphic genes near major repeats in intact males versus females. A contingency table with the two bins with positive values (distance= 0bp and distance= 100 Kbp-1Mbp) was setup to apply a chi-square test, which found the result significant at $p<0.05$ (Chi-square-test online calculator http://www.socscistatistics.com/tests/chisquare2/Default2.aspx). Bottom: Venn diagram (Oliveros, 2008) representation of the subset of DE genes located at 0bp from major satellite, highlighting that the 14 genes are all testosterone sensitive, where only one gene is specific to the comparison between intact XY male and GDX XY male.

In addition, the distances were also plotted based on three gene groups: “All genes”, corresponding to those in the reference genome assembly (mm9); “Expressed genes” obtained from the RNA-seq data (20 alignment reads per gene with two mismatched allowed – TopHat software); “Differentially expressed genes” (DESeq, p value< 0.05 of any fold change), obtained from comparing the transcriptomes from the four genotypes of interest from FCG as mentioned before in Figure 7 (Figure 8).

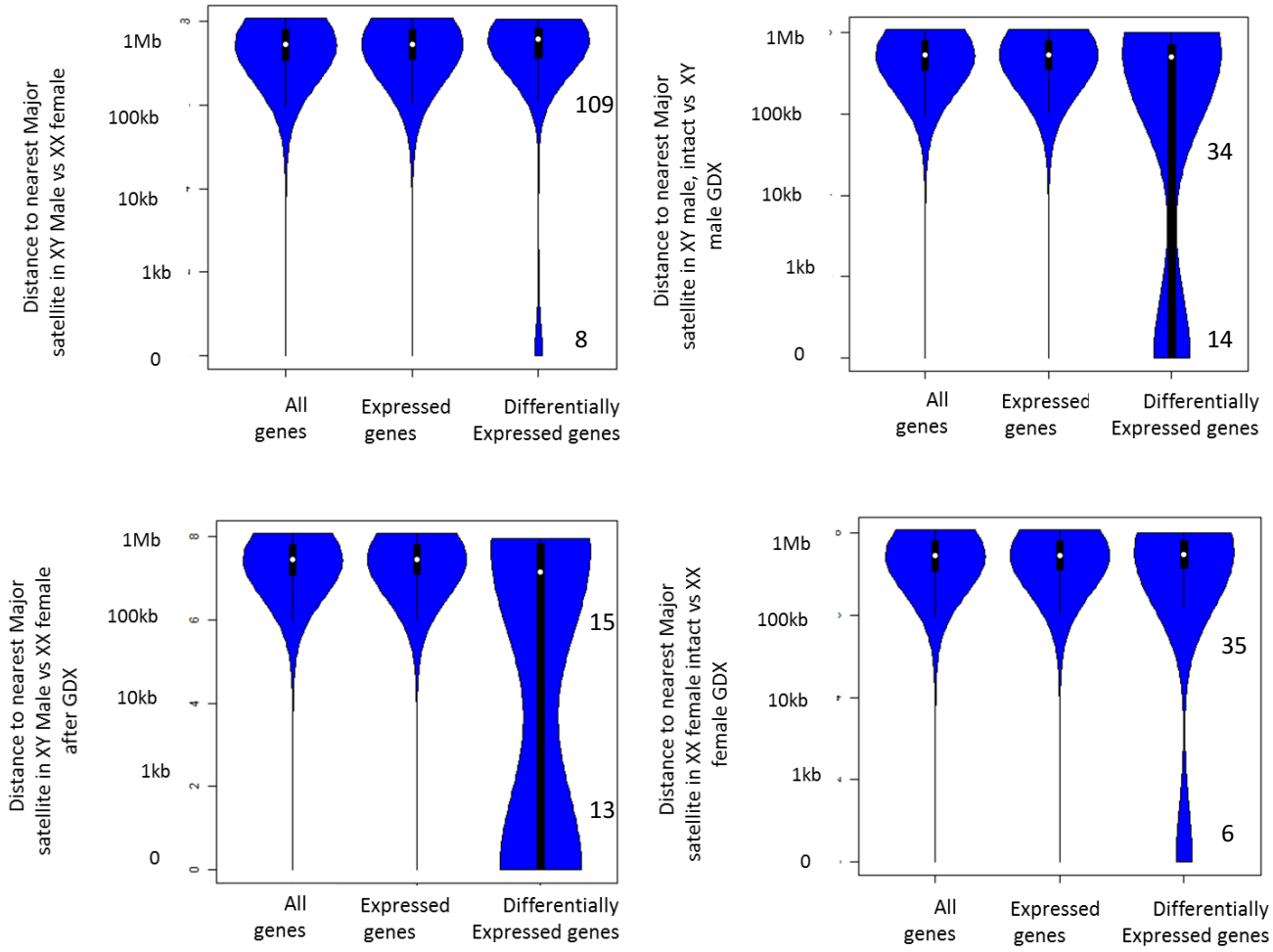


Figure 8 – The subset of differentially expressed genes in XY males and XX females (FCG) are located near major satellite repeats.

Violin plots of the distance of gene to the closest major satellite repeat (Y axis: $\log_{10}$ scale). Top left: compares XY males with XX females; Top right: compares XY intact males with XY gonadectomised males; Bottom left: compares XY gonadectomised males with XX gonadectomised females; Bottom right: compares XX intact females with XX gonadectomised females. All plots exhibit a subset of differentially expressed genes that are closer to repeats that is more prominent in the comparisons between gonadectomised male versus female and intact male versus gonadectomised male, further indicating the effects of gonadal hormones in potentially regulating and equalising major satellite repeat and repeat

proximal gene expression between the sexes. Note that the number of genes represented in each violin plot is also indicated in top plot from figure 7 with indicated regions in represented bins.

In addition to the distance from intergenic repeats (y axis), violin plots also show the probability density for the data set at individual values, which is proportional to the width of the coloured area of the plot. This analysis revealed a subset of differentially expressed genes that are positioned closer to repeats in intact versus gonadectomised XY males, and smaller subset in intact XY males versus intact XX females. Interestingly, all the DE genes identified were located in specific regions of chromosome 2 and 9 (Figure 9-A).

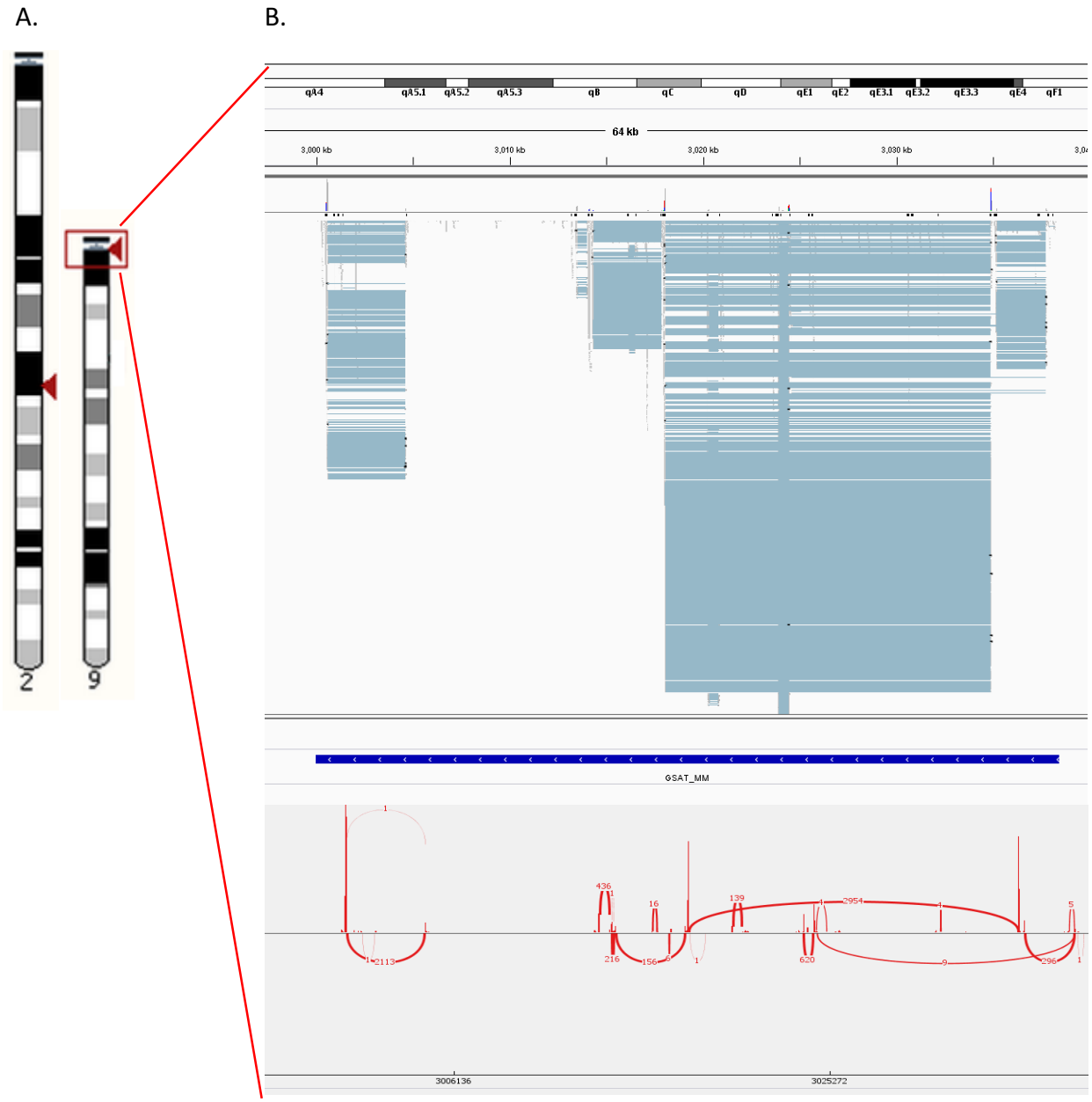
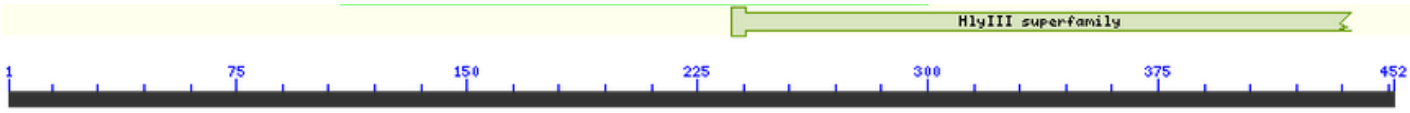


Figure 9 – DE genes near intergenic repeats location.

(A) Chromosome 2 and 9 location of DE genes near repeats (B) Visualization using the IGV browser of the block of satellite repeat on chromosome 9. From top to bottom, figure shows the location on the chromosome, reads on the bam file, location of the block of chromosome (marked in red in Table 2) and splicing of these reads.

Moreover, one of the hits located inside this repeat block (and DE in both cases presented) was the predicted gene Gm10722 (ENSMUSG00000091028); NCBI analysis tools for protein domain analysis (Marchler-Bauer et al., 2015) found conserved domains including the Haemolysin-III (Hly III) superfamily, of which members of this family are integral membrane proteins and are thought to possess progesterone binding regions in mammals (see discussion) (Figure 10, top) (Marchler-Bauer et al., 2015; Thomas et al., 2007). Interestingly, the major satellite monomer sequence could be mapped in both intronic and exon 3 of this gene (Figure 10, bottom).

A.



B.

```
>chromosome:GRCm38:9:3000996:3001827:-1
3001827 TCCACTTGACAACTTGAAAAATGACGAAATCACTGAAAAACGTGAAAAATGAGAAATGCA 3001768
3001767 CACTGTAGGACTTGGAATATGGCGAGAAAACTGAAAATCACGGAAAATGAGAAATTCACA 3001708
3001707 CTTTAGGACGTGAAATATGGCGAGGAAAACTGAAAAAGGTGGAAAATTTAAAAATTTCCA 3001648
3001647 CTCTAGGACATGGAATATGGCAAGAAAAATGAAAATCAAGAAAAATGAGAAACATCCACT 3001588
3001587 TCACAACTTGAAAAATGACAAAATCACTGAAAAACGTGAAAAATGAGAAATGTACACATA 3001528
3001527 GGACCTGGAATATGGCGAGAAAACTGAAAATCACGGAAAATGAGAAATACACACTTTAGG 3001468
3001467 ACGTGAAATATGGCGAGGAAAACTGAAAAAGGTGGAAAATTTAGAAATGTCCATTGTAGG 3001408
3001407 ACATGGAATATGGCAAGAAAACTGAAAATCATGGAAAATGAGAAACATCCACTTGACAAC 3001348
3001347 TTGAAAAATGACGAAATCACTGAAAAACGTGAAAAATGAGAAATGCACACTGTAGGACCT 3001288
3001287 GGAATATGGCGAGAAAACAGAAAATCACTGAAAATGAGAAATACACACTTTAGGACATGA 3001228
3001227 AATATGGCGAGGAAAACTGAAAAAGGTGGAAAATTTAGAAATGTCCACTGTAGGACATGG 3001168
3001167 AATATGGCAAGAAAACTGAAAATCATGGAAAATGAGAAACATGCAGTTGACAACTTGAAA 3001108
3001107 AATGACAAAATCACTGAAAAACGTGAAATATGAAAAATGCACACCGTAGGACCTGGAATC 3001048
3001047 TGGCGATAAAACTGAAAATCACGGAAAATGAGAAATACACACTTTAGGACGT          3000996
```

Figure 10 – Hit gene differentially expressed between male and females NCBI analysis.

The hit ENSMUSG00000091028, annotated as a protein-coding gene, had a homology to a conserved protein domain with a progesterone binding region in mammals (top) and in its exonic sequence (exon underlined in beige), part of it is made of major satellite monomer (coloured in red).

Another interesting feature is that a sequence with homology to the androgen response element sequence is present within the major satellite monomer, highlighted below, where blue characters represent perfect match and in red, unpaired bases.

Major satellite monomer:

GGACCTGGAATATGGCGAGAAAACTGAAAATCACGGAAAATGAGAAATACACACTTTAGGACGTGAAATATGG
CGAGGAAAACTGAAAAAGGTGGAAAATTTAGAAATGTCCACTGTAGGACGTGGAATATGGCAAGAAAACTGA
AAATCATGGAAAATGAGAAACATCCACTTGACGACTTGAAAAATGACGAAATCACTAAAAAACGTGAAAAATG
AGAAATGCACACTGAA

What contribution to the overall major satellite expression is made by these novel genes?

In view of the fact that there was overlap between the repeat sequence at exons it was possible that the reads obtained from the RNA seq data were coming from major satellite repeat sequence or exons or both. In order to unambiguously distinguish between these possibilities, a new R script pipeline was devised with Dr Thomas Carrol (MRC-CSC) and was employed to exclude the exonic regions and compare to the total reads. The results are presented in figure 11 and show that the non-exon major satellite repeat is responsible for the bulk of the expression.

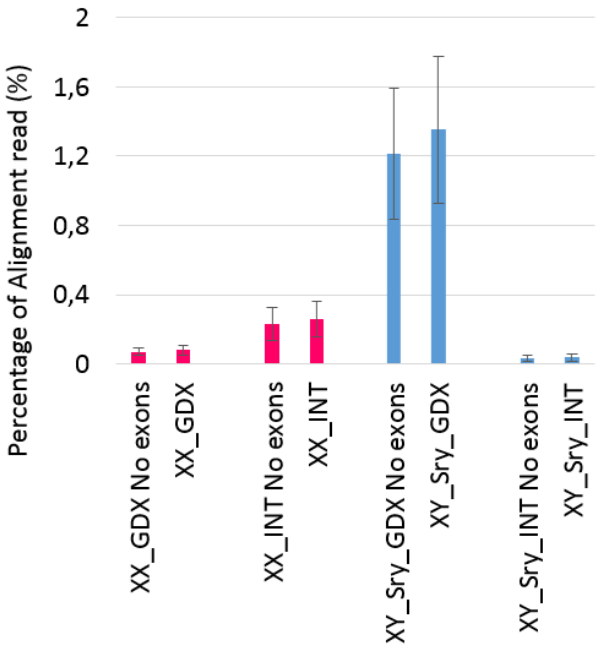


Figure 11 – Repeat expression of the 42 intergenic major satellites on XX females and XY males (FCG mice) is not affected by exonic regions of nearby genes with major satellite monomer codified in their exons.

Removal of exonic regions is noted in the plot as "No exons", and major satellite expression is compared for XX intact females (XX_INT) and for XY males intact and gonadectomised (INT and GDX, respectively). N=3 biological replicas.

Considering the major satellite expression in the different genotypes, the level of expression of the major satellite proximal DE genes was assessed on the relevant genotypes for comparison (Figure 12).

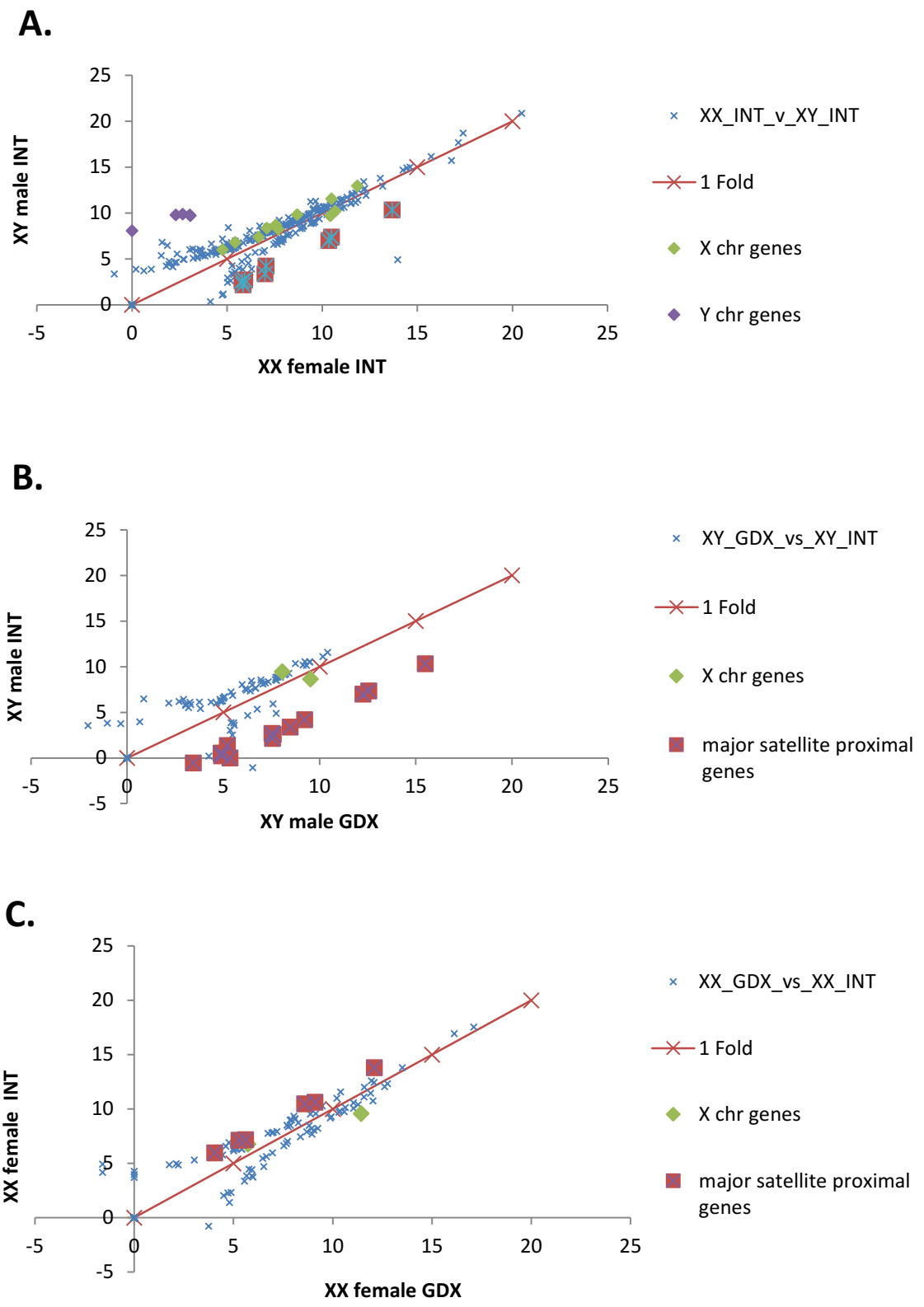


Figure 12 – The major satellite proximal DE genes correlate with major satellite expression and have a trend for highest expression.

Expression of DE genes (log2) is presented in plots (A), (B) and (C). Of those DE genes, X chromosome genes are in green, Y chromosome genes are in purple, and major satellite proximal genes are represented in brown squares. (A) compares intact XX female with intact XY male: (B) compares intact XY male with GDX

XY male; and (C) compares intact XX female with GDX XX female. In the 3 cases, the major satellite proximal DE genes, are amongst the most affected DE genes.

Interestingly, major satellite expression correlates with expression of the major satellite proximal genes, and these were in fact amongst the genes that had highest expression level and differed the most in the comparisons shown.

Overall the results suggest an interesting model involving major-satellite repeat elements, testosterone and the regulation of major satellite repeat-proximal genes whereby the repeats are potentially regulating the hormone response of these genes.

3.3- Heterochromatic foci distribution in gonadectomised male mice

In view of the fact that gonadectomy lead to increased major satellite repeat expression this raised the possibility that heterochromatin might be compromised as has been previously suggested in BRAC1 mutations (Zhu et al., 2011). There may be a number of different effects of gonadectomy on the heterochromatic foci. For example one might expect that expression of the heterochromatin repeats would loosen their association and result in a change in the number of heterochromatic foci. Therefore, to further understand the effects of gonadectomy in our XY males, heterochromatin structure was assessed by developing a novel methodology using Imagestream which combines microscopy with flow cytometry allowing quantification of large numbers of cells. This analysis was done on blood samples from transgenic male mice expressing GFP fused to HP1-β in T-cells (Fig 11), as HP1β foci localize to chromatin dense regions (Festenstein et al., 2003). This approach allowed visualisation of the heterochromatic foci in living cells and aimed to assess the stability of the chromatin structure in the absence of testosterone.

Blood samples were collected before and after gonadectomy, over a period of 12 weeks (Week 0, Week2, Week 6 and Week 12), to assess the distribution of the number of heterochromatic foci per cell. The base-line measurement Week 0, was prior to any surgical procedure. Five mice were gonadectomised at 10 weeks of age by removing the testis, and 5 control mice were also subjected to the same procedure but without removal of the gonads (sham).

Ex vivo peripheral blood mononuclear cells (PBMCs) were isolated and stained with Draq5^{TM}, a far-red DNA stain that is suitable for living cells and less toxic than DAPI (Kang et al., 2011), and

that similarly stains chromatin dense regions. Because the HP1β-GFP proteins are only expressed in T lymphocytes, it was possible to distinguish these cells from other PBMCs.

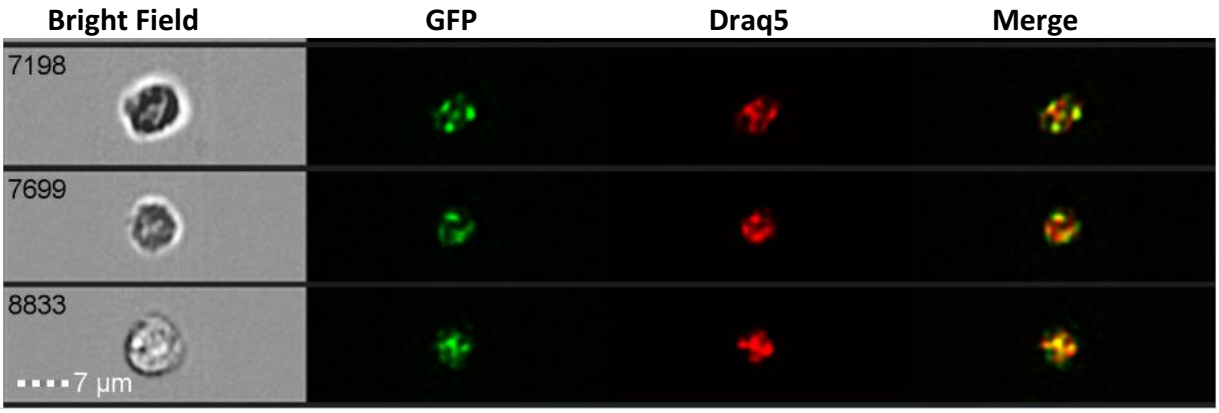
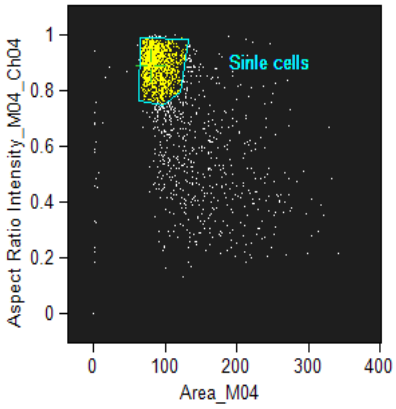


Figure 13 - Peripheral mouse T –cell imaging using the ImageStream X.

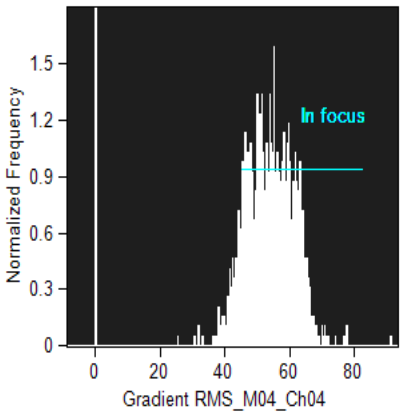
The green shows HP1β-GFP, red Draq5™ staining for nucleus delimitation and heterochromatic foci, yellow represents the merge between GFF and Draq5 colours.

The analysis included heterochromatic foci counting: (i) GFF spots on T-cells; (ii) Draq5 spots on total PBMCs; (iii) Draq5 spots excluding GFP positive cells (T-cells); and, (iv) Draq5 spots on T cells. The tight analytical single-cell gate excluded large cells, doublets, and cell debris (Figure 14-A and 14-E); a second gate for granularity excluded granular cells, out of focus cells and apoptotic cells but not lymphocytes (Figure 14-B and 14-F). Finally cells are gated according to GFP or Draq5™ fluorescence (Figure 14-C and 14-G) and plotted according to the number of spots per cell (using Ideas Software tools). The gated cells are therefore mainly T and B lymphocytes, but it cannot be excluded that other remaining PBMCs were also present however, these would only represent less than 10% (supplementary 6.2).

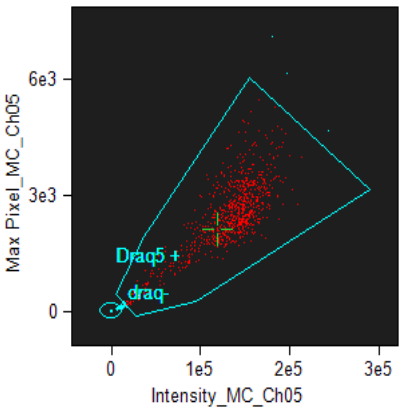
A.


Sinle cells
Aspect Ratio Intensity_M04_Ch04
Area_M04
0
0.2
0.4
0.6
0.8
1
0
100
200
300
400

B.


In focus
Normalized Frequency
Gradient RMS_M04_Ch04
0
0.3
0.6
0.9
1.2
1.5
0
20
40
60
80

C.


Draq5 +
draq-
Max Pixel_MC_Ch05
Intensity_MC_Ch05
0
3e3
6e3
0
1e5
2e5
3e5

D.


Ch04
Ch02
Ch05
Ch02/Ch05
54
222
600
748
950
960
990
1091
1207
1320
1797
1843
1910
7 µm

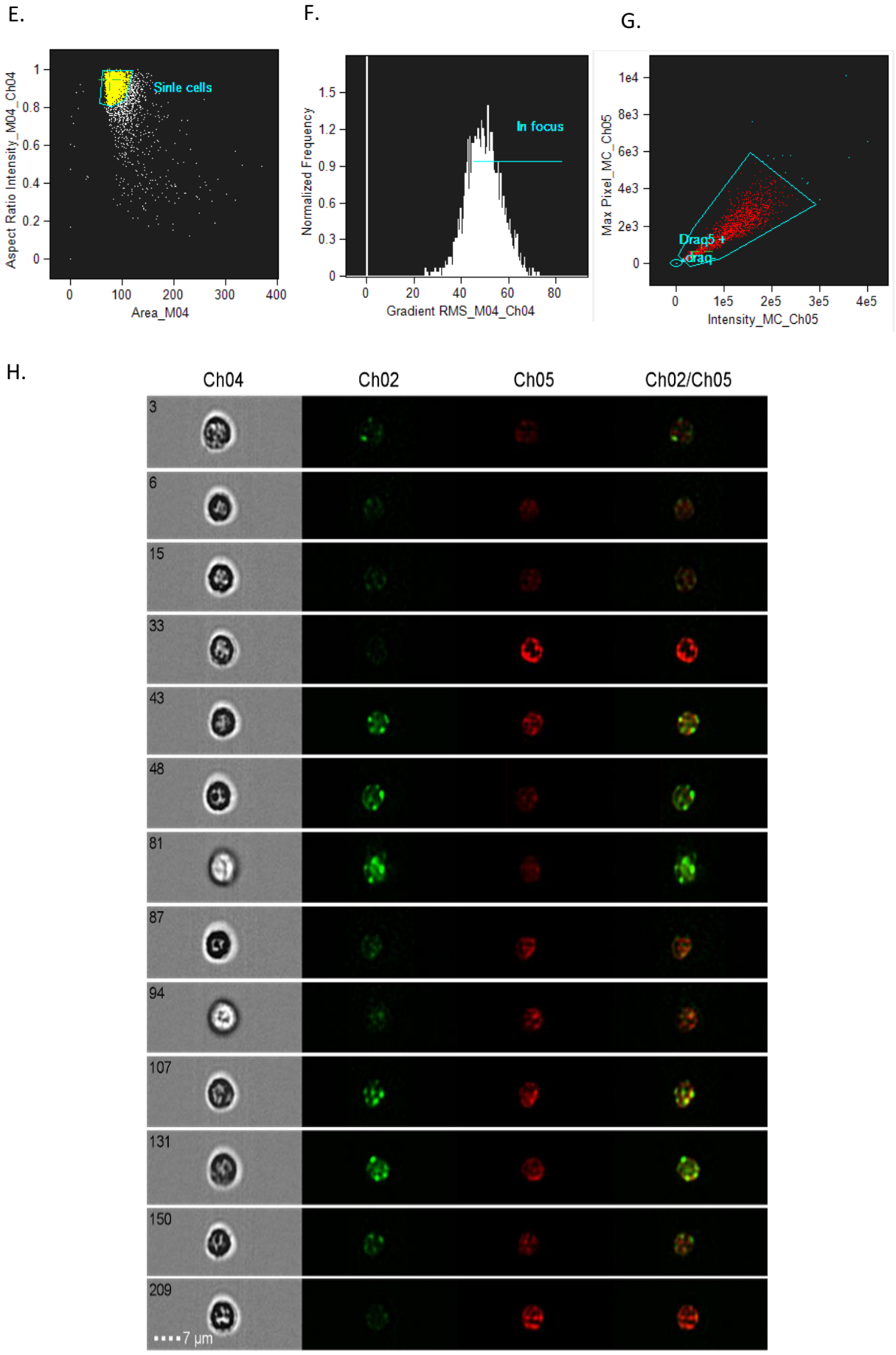


Figure 14 – Ideas Software analysis of heterochromatic foci.

Mouse 6127 sham control (A, B, C, D) and mouse 6114 gonadectomised (E, F, G, H) at week 12 post surgery. Channel (Ch) 2 - GFP; Ch04- Bright Field (BF); Ch05- Draq5; **(A, E)** Single cell gate, plotting area of BF versus Aspect Ratio (estimate of cell size) in BF; **(B, F)** focussed cell gate, plotting the granularity of cells; **(C, G)** Draq5 positive cells gate, plotting the stain intensity; **(D, H)** Lymphocyte microscopy images, where the digits indicate the event number, and the scale is shown at the end of each list.

Representative data is shown in figure 14 which allows us to compare data from a gonadectomised mouse with that from an intact mouse. Cell morphology and Draq5 staining are also homogenous and consistent across both groups.

GFP positive cell analysis

Cells collected in this analysis are T-cells, as the transgene for GFP-HP1β is specific for this cell type (Festenstein, Pagakis et al. 2003). According to the sample size (Figure 15-A), the first measurement ("Week 0"–W0) had the highest amount of cells. Also on the base-line measurement (W0), both groups presented the same number of clusters per cell distribution (Figure 15-B). The respective distributions of number of centromeric clusters per cell over time is presented in Figure 13. Gonadectomised mice (16-A) have a very similar distribution compared with sham mice (16-B). This control group in particular, shows a close overlap of distributions across 12 weeks, which confirms the robustness of the technique. Although there was a statistically significant difference from week 2 to week 6 on gonadectomised mice (t-test, p-value < 0.001), week 12, showed no significant difference with remaining weeks. Overall, gonadectomy does not seem to affect the number of HP1β containing clusters per T cell.

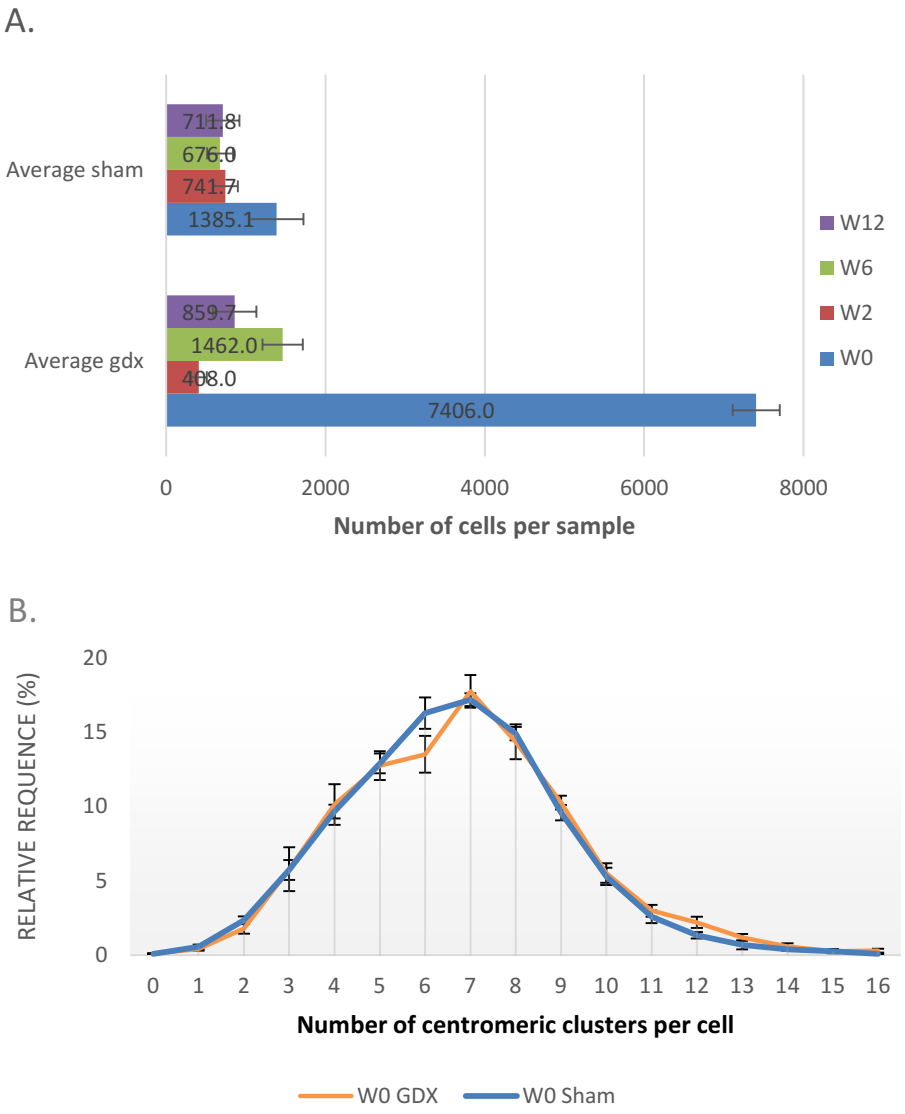


Figure 15 – Acquisition of T cells selected from mice blood with ImageStreamX, before and after gonadectomy.

The number of cells in each group from each week is presented in (A). The base-line measurement at week 0 (W0) is presented in (B); this plot shows no significant alteration in both groups**.** Blood samples from gonadectomised (in orange) and sham (in blue) male mice were analysed using ImageStreamX. Between 300 and 8000 T-cells were considered. Statistical significance was assessed via t-test (unpaired, 3 tails). N=5 biological replicas; SEM=error bars.

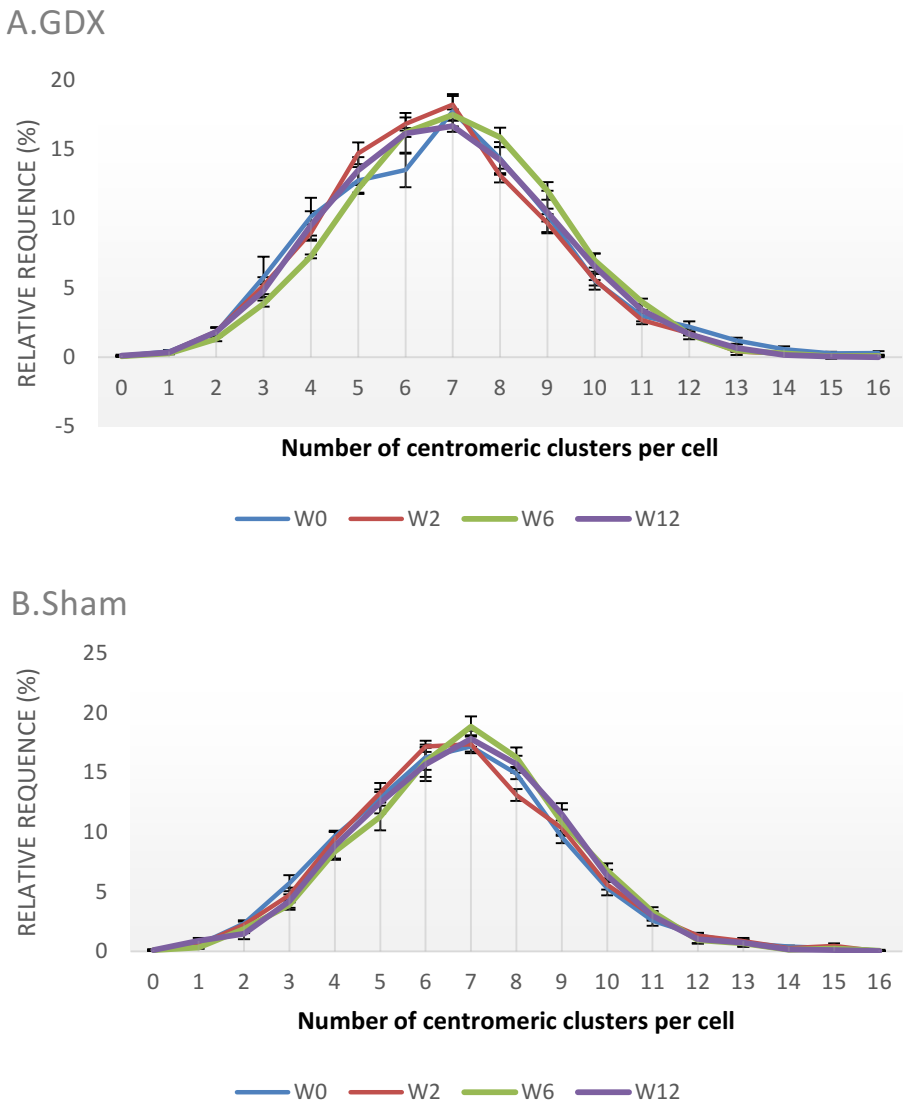


Figure 16 – HP1β-centromeric cluster distribution from blood T cells, before and after gonadectomy.

(A) Gonadectomised mice showed no effect on HP1β clusters after gonadectomy; (B) sham mice HP1β-cluster distribution remains unchanged across blood measurements. Blood samples from gonadectomised (A) and sham (B) male mice were analysed using ImageStreamX. Between 300 and 8000 T-cells were considered. Statistical significance was assessed via t-test (unpaired, 3 tails). N=5 biological replicas; SEM=error bars.

DRAQ5™ positive cells analysis

PBMCs selected after red cell lysis were also stained with DRAQ5™ which allowed us to track DNA rich regions in PBMCs other than T-cells. As mentioned given the gates, the cells analysed are mainly B and T cells. In Figure 17 the base-line measurement (W0) on the other hand, showed a difference in the 2 randomly selected groups of mice which were to be subjected to gonadectomy or sham. No obvious reason was found for this difference.

The respective distributions of heterochromatic foci per cell over time is presented in Figure 18. Gonadectomised mice (18-A) have a steady change over time with the curve shifting to the right – indicating an increase in the number of heterochromatic foci per cell. On the other hand, Sham mice (18-B) present a very similar and overlapping distribution on weeks 0, 2 and 6, and a shift to the right only on week 12, although this was not statistically significant.

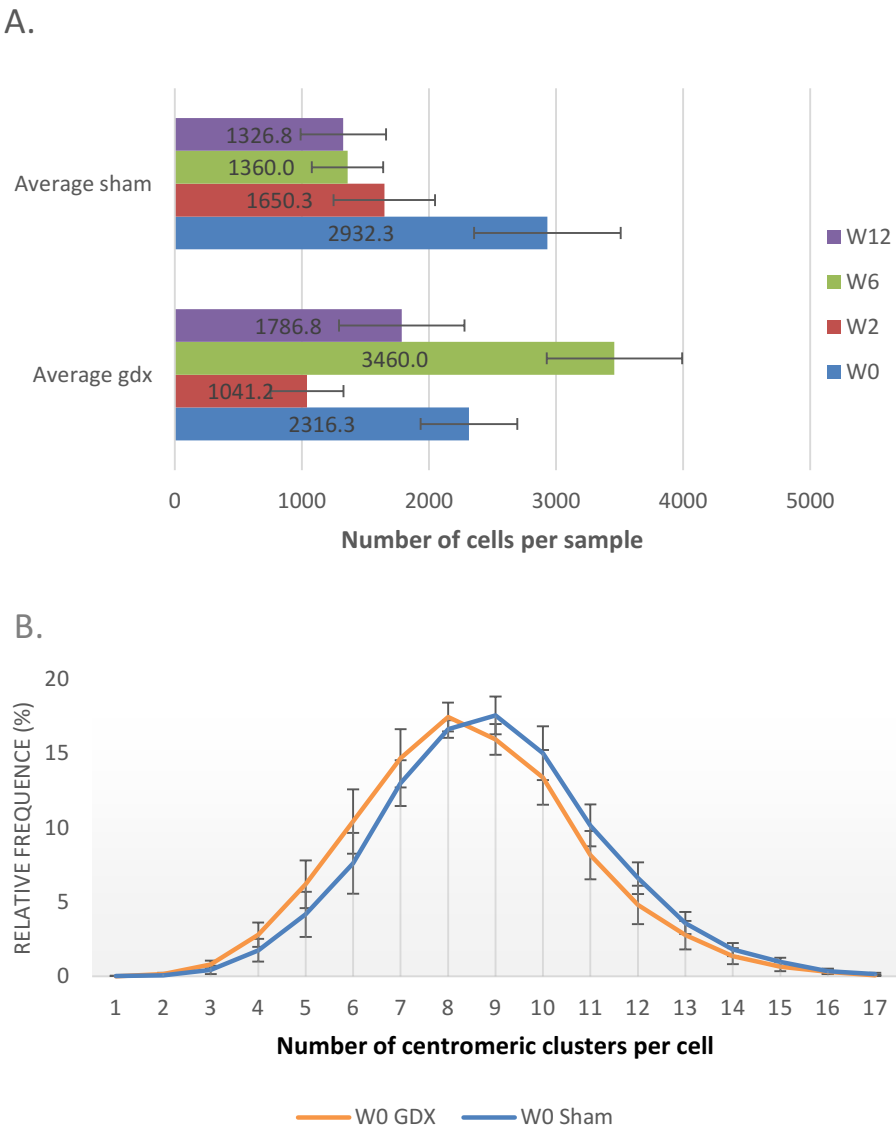


Figure 17 – Acquisition of T and B cells selected from mice blood with ImageStream, before and after gonadectomy (Draq5™ staining).

The number of cells in each group from each week is presented in (A). The base-line measurement at week 0 (W0) is presented in (B); this plot shows a slight shift of the distribution of Sham mice to the right. Blood samples from gonadectomised (in orange) and sham (in blue) male mice were analysed using ImageStreamX. Between 800 and 1200 T-cells per blood sample were considered. Statistical significance was assessed via t-test (unpaired, 3 tails). N=5 biological replicas; SEM=error bars.

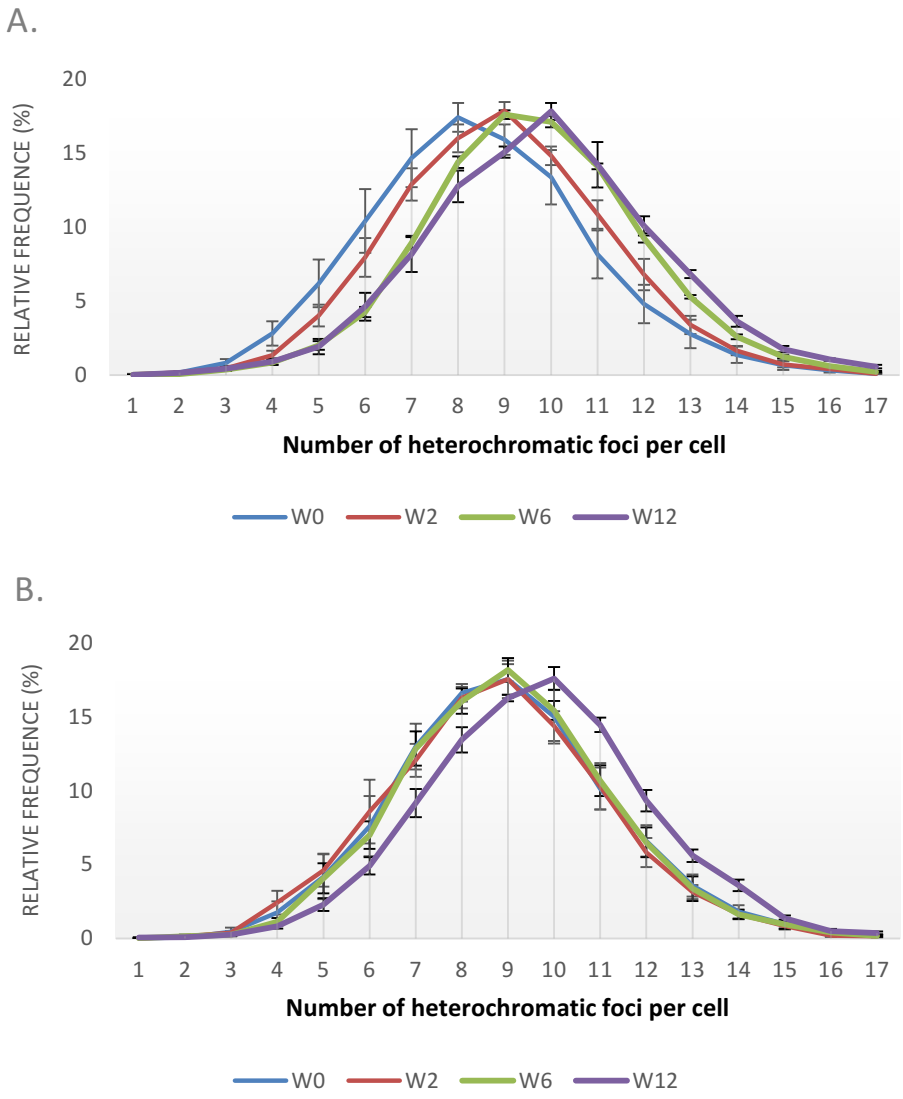


Figure 18 - Heterochromatic foci distribution from PBMCs, before and after gonadectomy.

(A) Gonadectomised mice showed a steady increase of the number of clusters per cell after gonadectomy; (B) sham mice cluster distribution remains unchanged except on week 12, where the number of spots increases per cell. Comparing both groups on W12, the distributions show no difference. Blood samples from gonadectomised (A) and sham (B) male mice were analysed using ImageStreamX. Between 300 and 8000 T-cells were considered. Statistical significance was assessed via t-test (unpaired, 3 tails). N=5 biological replicas; SEM=error bars.

DRAQ5™ positive+ GFP negative cells analysis

T cells were gated out of the sample by GFP fluorescence. By deduction the final population would be composed of mainly B cells. The percentage of B cells in the blood is smaller than T-cells (Supplementary 7.2), and as expected the sample size is smaller for these populations (Figure 19-

A). Consequently, the variation among individuals would be expected to be bigger, consistent with the larger error bars. The base-line measurement (W0) shows little difference between both groups.

The respective distributions of heterochromatic foci per cell over time is presented in Figure 20. Similar to total Draq5, gonadectomised mice (20-A) have a steady increase in number of foci per cell over time, and sham mice (20-B) present a very similar and overlapping distribution on weeks 0, 2 and 6, and a shift to the right on week 12, although not statistically significant.

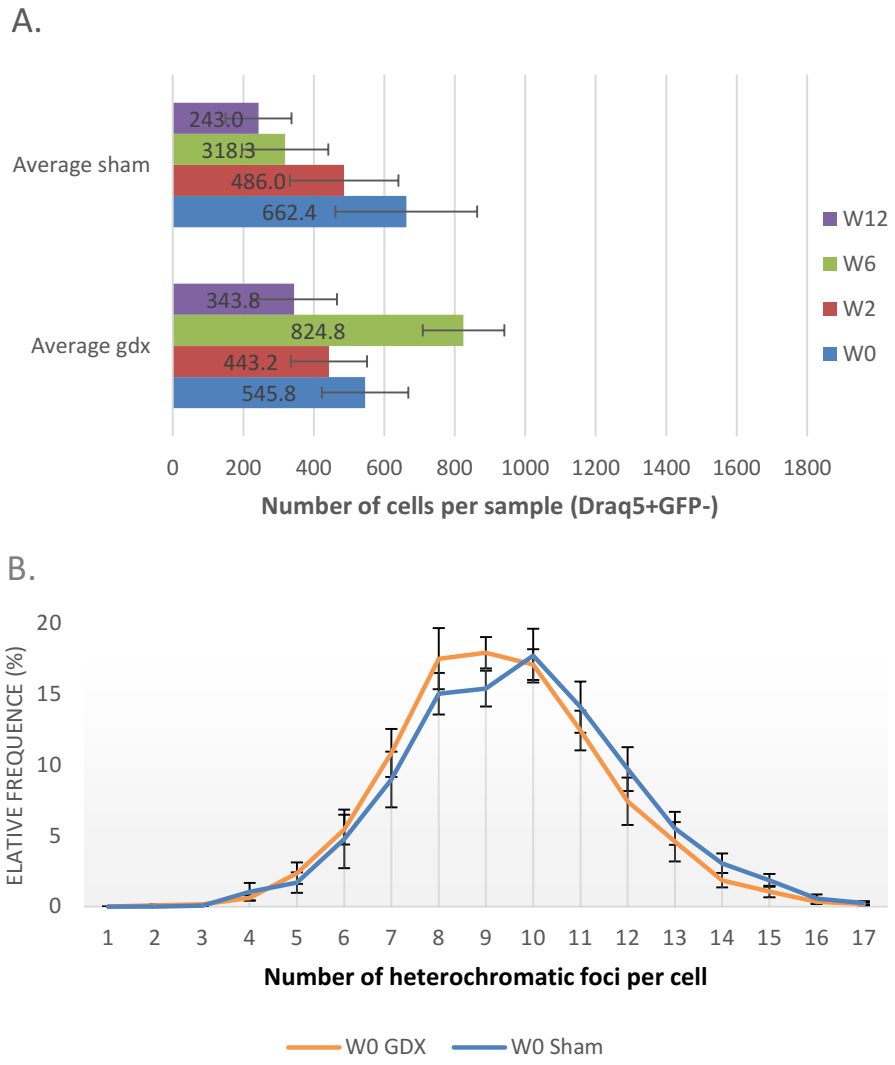


Figure 19 - Acquisition of PBMCs except T cells with ImageStreamX, before and after gonadectomy (Draq5™ staining).

The number of cells in each group from each week is presented in (A). The base-line measurement at week 0 (W0) is presented in (B); this plot shows a very slight shift of the distribution of Sham mice to the right,

that due to sample variation is not significant. Blood samples from gonadectomised (in orange) and sham (in blue) male mice were analysed using ImageStreamX. Between 200 and 900 T-cells per blood sample were considered. Statistical significance was assessed via t-test (unpaired, 3 tails). N=5 biological replicas; SEM=error bars.

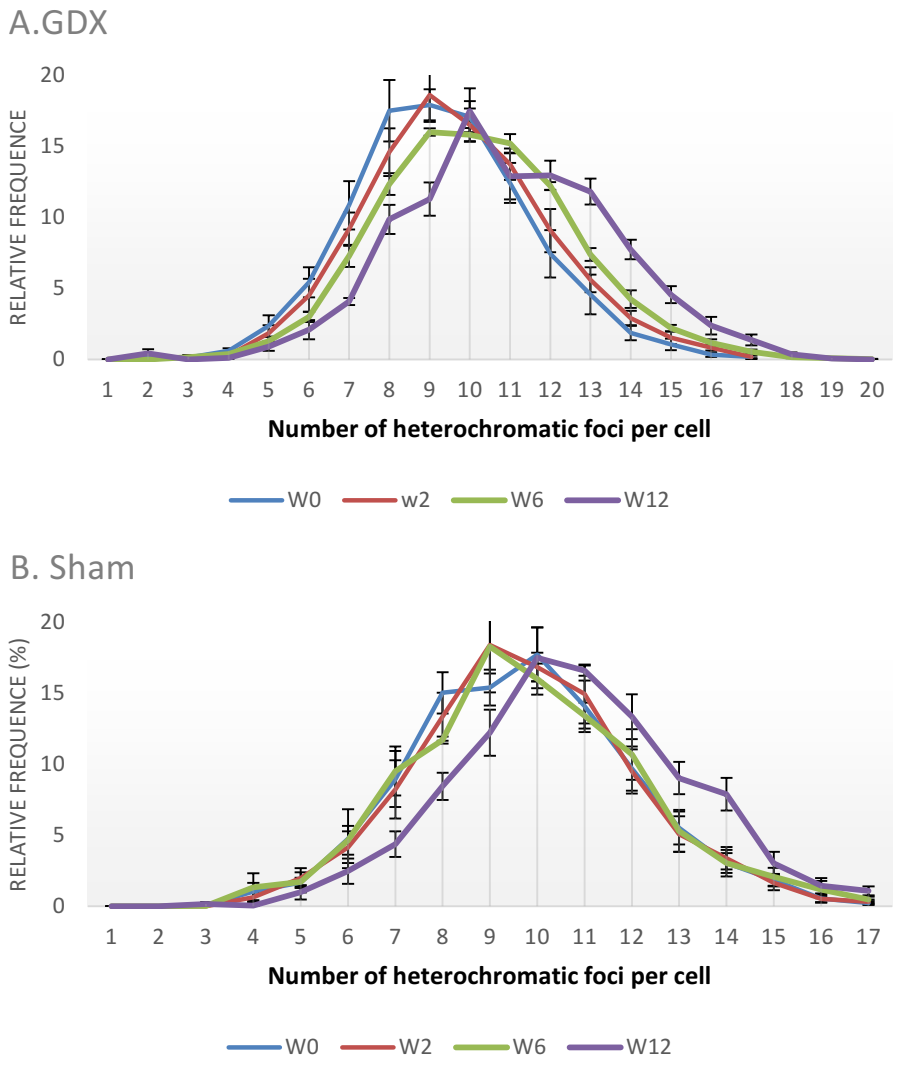


Figure 20 - Heterochromatic foci distribution from PBMCs excluding T-cells, before and after gonadectomy.

(A) Gonadectomised mice showed a steady increase of the number of clusters per cell after gonadectomy; (B) sham mice cluster distribution remains unchanged except on week 12, where the number of spots increases per cell. Comparing both groups on W12, the distribution shows no difference. Blood samples from gonadectomised (A) and sham (B) male mice were analysed using ImageStreamX. Between 200 and 900 cells were considered. Statistical significance was assessed via t-test (unpaired, 3 tails). N=5 biological replicas; SEM=error bars.

DRAQ5™ positive+ GFP positive cells analysis

In this section, the opposite stratagem was applied: GFP positive T cells, were selected for Draq5 heterochromatin foci analysis. Figure 21 represents the number of cells per blood sample, which is very similar to Figure 17-B. In theory, all T-cells should be stained with Draq5, but due to fluorescence threshold, some cells could have been excluded. As for the other Draq5 stained populations, the base-line measurement (W0) showed a small difference between both groups, where sham mice have an increment of 1 spot per cell. This base-line difference was not present in the HP1β-GFP cluster measurement.

The respective distributions of heterochromatic foci per cell over time is presented in Figure 22. Similar to total Draq5, gonadectomised mice (17-A) have a steady change over time to the right, and sham mice (17-B) present a similar distribution on weeks 0, 2, 6 and 12. However, unlike the other Draq5 populations, there was no shift to the right on week 12 in the sham mouse group.

Statistical analysis showed that the distribution of gonadectomised mice from week 0 and week 2 compared to week 12 was significant (t-test, p-value < 0.02).

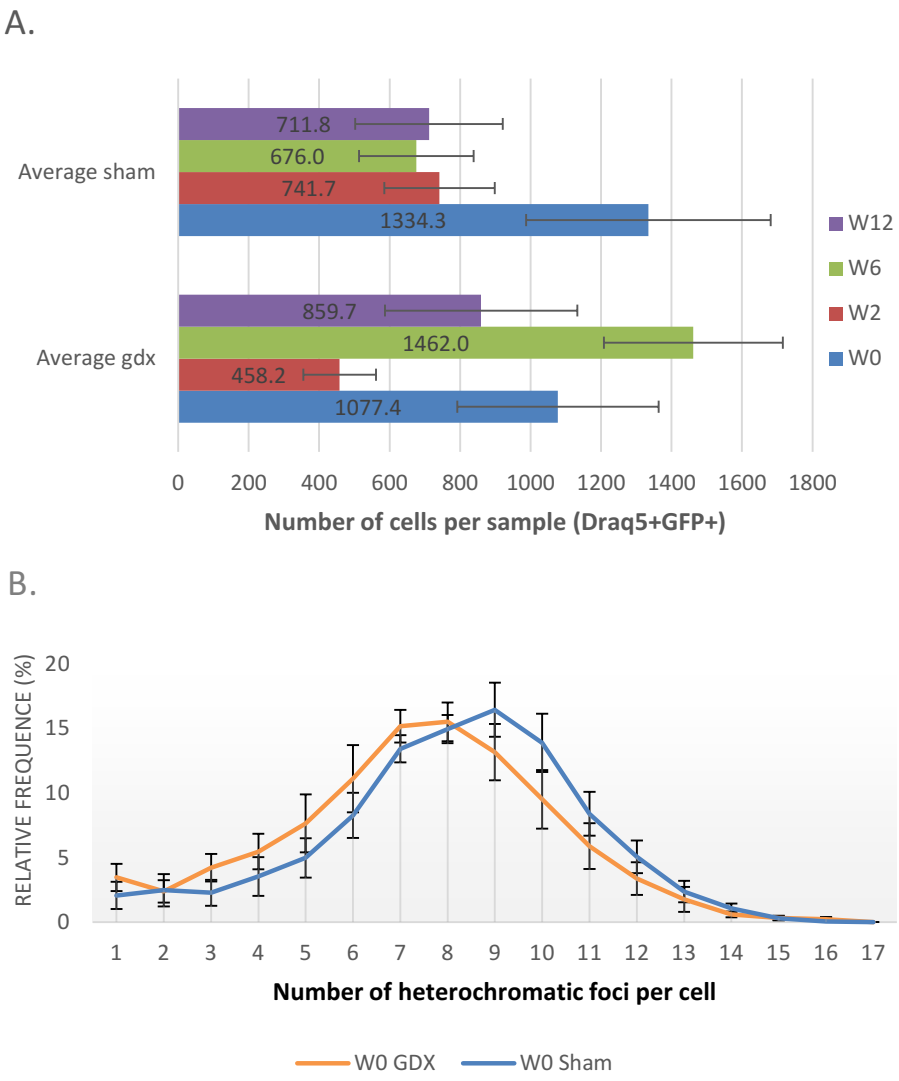


Figure 21 - Acquisition of T cells with ImageStream, before and after gonadectomy (Draq5™ staining).

The number of cells in each group from each week is presented in (A). The base-line measurement at week 0 (W0) is presented in (B); this plot shows a very slight shift of the distribution of Sham mice to the right, that due to sample variation is not significant. Blood samples from gonadectomised (in orange) and sham (in blue) male mice were analysed using ImageStreamX. Between 400 and 1900 T-cells per blood sample were considered. Statistical significance was assessed via t-test (unpaired, 3 tails). N=5 biological replicas; SEM=error bars.

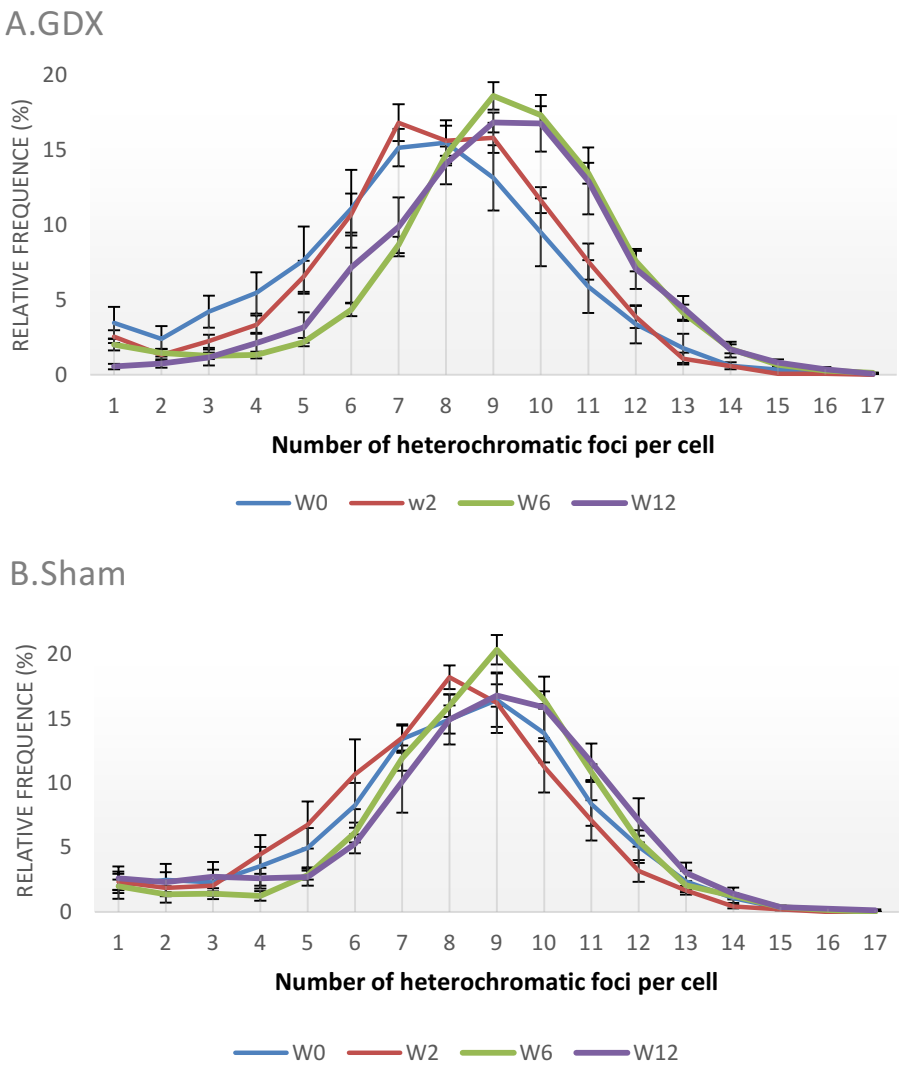


Figure 22 - Heterochromatic foci distribution of T-cells, before and after gonadectomy.

(A) Gonadectomised mice showed a steady increase of the number of clusters per cell after gonadectomy; (B) sham mice cluster distribution remains overall unchanged. Comparing both groups at W12, the distributions of gonadectomised mice showed an increment of 1 spot per cell. Blood samples from gonadectomised (A) and sham (B) male mice were analysed using ImageStreamX. Between 400 and 1800 cells were considered. Statistical significance was assessed via t-test (unpaired, 3 tails). N=5 biological replicas; SEM=error bars.

Overall, due to differences in base-line measurements and the shift of sham samples on week 12, it is difficult to make definitive conclusions. Further time points and tissue analysis would be required. In summary, the methodology developed was robust at identifying the HP1-GFP spots as long as the sample size provided was adequate. These results showed no change in the number of heterochromatic foci per cell after gonadectomy. However, the Draq5 analysis showed an

intriguing trend towards increasing number of foci per cell following gonadectomy. This result suggests that the two methods might be identifying different heterochromatin forms (see discussion 4.2)

3.4- Effect of testosterone in T-cells *in vitro*

From the list of Y encoded genes, a few candidates were identified as potential participants in the Y-chromosome-SRY-testosterone mediated heterochromatic silencing. These include the histone demethylase KDM5D that is responsible for removing the active marks H3K4me3/2, therefore potentially promoting silencing. Given the critical amount of data in T cells, it was attempted to set up an in vitro model to study these candidates.

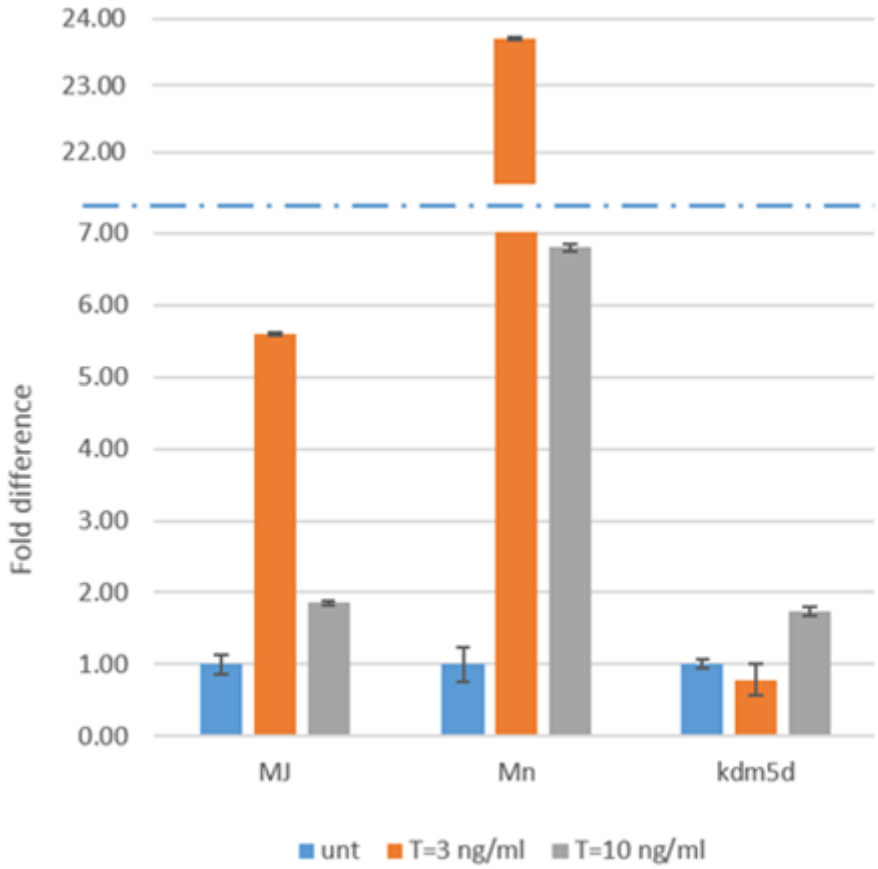


Figure 23 - Expression levels of Kdm5d mRNA and heterochromatic repeats in activated T-cells.

Lymph nodes were collected from 6 weeks old C57Bl6 mice; B-cells were pooled and depleted from sample. T-cells were activated with Il2 and magnetic beads for 9 days in culture. Expression values are relative to untreated cells percentage levels. Testosterone treatment lasted 3 days (3ng/ml, orange and 10ng/ml, grey). N=3 technical replicas (statistical analysis was not possible); SEM=error bars.

Due to the technical difficulties of culturing thymocytes, T-cells were cultured from lymph nodes. From a pool of 5 wild type male mice, T cells were selected using magnetic beads and cultured in a hormone free medium. Since lymphocytes in this tissue are resting (G0), cells were activated by IL-2 (interleukin-2) and magnetic beads coupled to anti-TCR and CD28 antibodies. After proliferation, cells were treated for 3 days at different testosterone concentrations (0, 3 ng/ml and 10 ng/ml) and collected for RNA processing. Next the heterochromatic silencing was assessed via qRT-PCR (Figure 23).

Upon testosterone treatment, T-cells upregulated both major and minor satellite repeat expression. Surprisingly, the lower concentration of testosterone exerted a higher effect on major and minor repeats. Such an effect was opposite to that seen in the ex vivo samples and might represent a partial agonist effect of testosterone, in addition, KDM5D responded in an opposite manner to the repeats (see discussion).

3.5- X–chromosome pair affects X linked genes expression & tandem heterochromatic repeats expression

From the genes identified as X-inactivation escapees, several are known to be involved in gene regulation at the chromatin level, including the histone demethylases KDM5C (aka, jarid1c) and Kdm6a (aka, Utx) - interestingly, these demethylases have homologs on the Y chromosome, Kdm5d (Jarid1d) (see above) and Kdm6b (Uty), respectively. Kdm5c differential expression between the genders was confirmed by qRT-PCR in thymus (Figure 24).

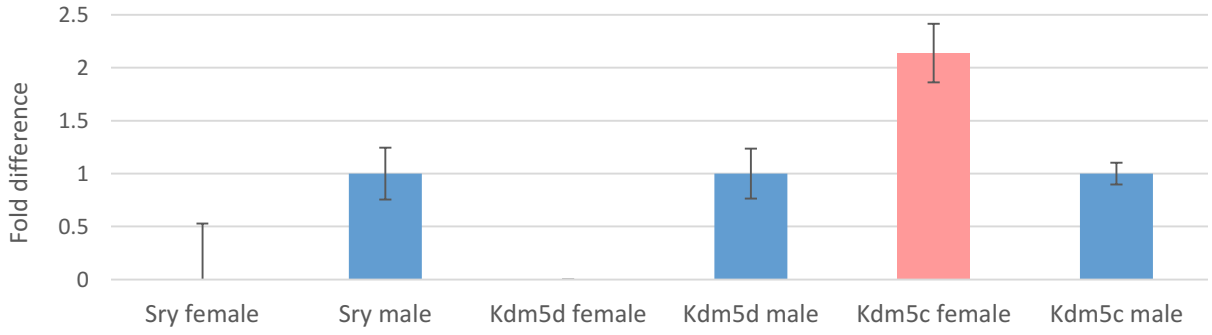


Figure 24- Sex chromosome linked mRNA expression in males vs. females.

Transcription of Y-encoded Sry and Kdm5d, and the X-encoded Kdm5c were assessed via qRT-PCR. Expression was normalized to the housekeeping gene GAPDH, a one male was used to normalize remaining

individuals. The plots shows Kdm5d is only expressed in XY background, whereas Kdm5c is expressed at 2 fold higher in XX females, confirming Kdm5c escapes X inactivation in female thymus. The Sry primer revealed some unspecificity for female tissue. N=3 technical replicas (statistical analysis was not possible); Error bars=SEM.

To further investigate whether the X chromosome dosage was influencing heterochromatin expression, either by providing a sink for heterochromatin factors (see discussion) or through increased expression of escapees from x-inactivation, mutant mice with differing numbers of X chromosome were studied. Thus, using mice with the Y chromosome attached to the X chromosome (Wijchers et al. 2010), it was possible to produce XO female and XXY male mice. Major and minor satellite expression were analysed in thymus and lymph nodes from these mice (Figure 25).

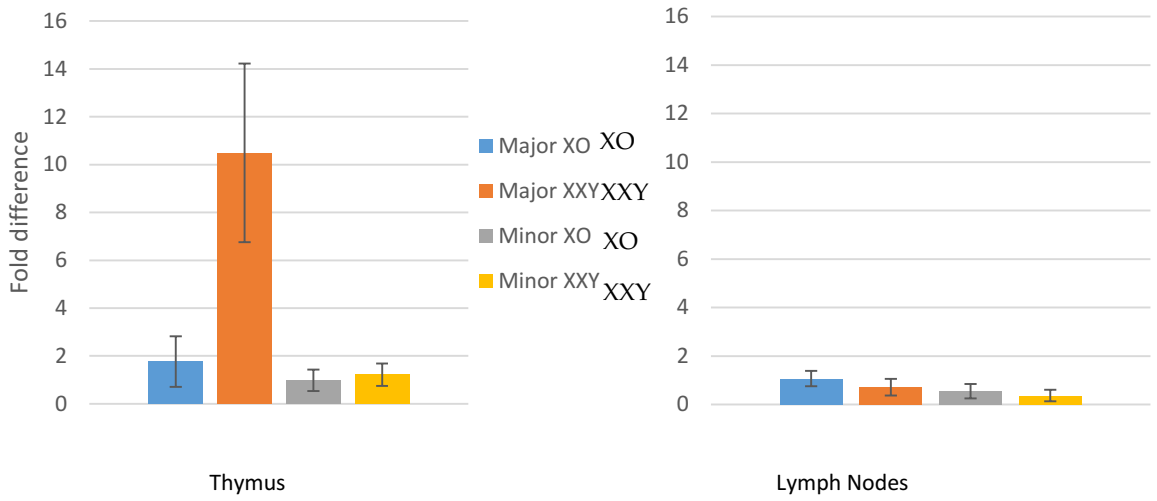


Figure 25 - Expression levels of major and minor satellite repeats in thymus and lymph nodes from XO and XXY mice.

Tissues were collected from four 8 weeks old albino mice. Expression levels were normalized to *GAPDH* housekeeping gene and to one of the XO replicas for comparison. N=4 biological replicas; SEM=error bars. T-student test revealed no statistical significance.

In thymocytes, major satellite expression was markedly upregulated in XXY males compared to XO female, whereas minor satellite showed no effect; on lymph nodes on the other hand, minor satellite seem to be slightly downregulated in XXY male compared to XO females, although not statistically significant. This result supports the hypothesis that increasing X chromosome dosage results in de-repression of major satellite repeats in thymus and that this effect is dominant over any Y chromosome effect in these mutant mice this is consistent with the effect on hCD2 variegation (Wijchers et al 2010).

3.6- Influence of age and gender in repeat expression of wild type males and females

Given the differences in age in mice from section 3.1 and 3.5, the expression of tandem satellite repeats in 8, 24 and 42 weeks old mice was assessed with qRT-PCR. This experiment also aimed at comparing males and females to assess if the Y-chromosome-testosterone effects were present in *wild type* mice. Two tissues were collected, thymus and lymph nodes from 4 biological replicas in each group. Besides RNA, tissue samples were also used for visualization of heterochromatin structure (Draq5^{TM} staining) with ImageStreamX (Figures 26 and 27).

In lymph nodes, there was no statistically significant difference in heterochromatic major satellite repeats expression between males and females (Figure 27-A,B). On the other hand, in thymocytes at weeks 8 and 24 weeks of age, major satellite repeat expression was higher in females consistent with the RNA seq data presented earlier, but at 42 weeks of age, curiously the males expression was higher than the females, although none of these differences reached statistical significance (Figures 26-A,B), see discussion. Moreover, female thymocytes3 showed an age dependency from mice 42 weeks old to 24 and to 8, and on male thymocytes, from 42 weeks old mice to 8 weeks old mice, major satellite expression was significantly different (t-test, p-value < 0.05).

The heterochromatin foci analysis is concordant with the expression data from major satellites. On the one hand, lymph nodes show no significant difference between age groups or gender (figures 27-C,D and 28); and on the other hand, the same comparisons that were significantly different for major satellite expression in thymocytes, were also significantly different in their heterochromatin foci distribution, but not for gender (Figure 26-C,D and 28). This suggests a positive correlation between major satellite expression and the number of heterochromatin foci per cell.

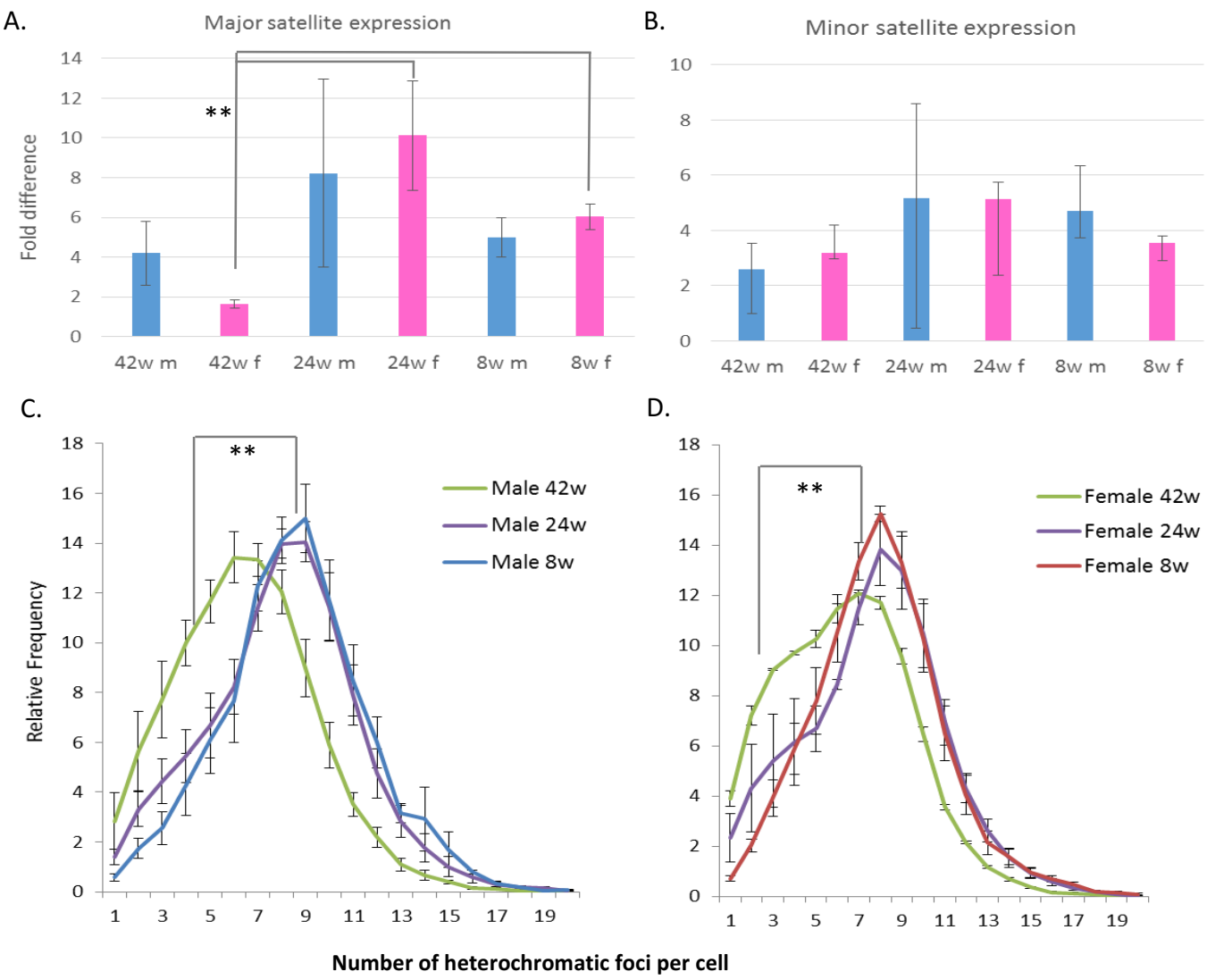


Figure 26 – Influence of age in chromatin structure and tandem repeat expression profile in thymus.

Major satellite (A) and minor satellite (B) repeat expression in thymocytes of males and females at ages 8, 24 and 42 weeks. Heterochromatin foci distribution in males (C) and females (D) of the same age. Thymocytes were stained with Draq5TM and analysed using ImageStreamX. Between 900 and 8000 cells were considered. Statistical significance was assessed via t-test (unpaired, 3 tails). N=3 or 4 biological replicas; SEM=error bars. N=3-4 biological replicas; SEM=error bars.

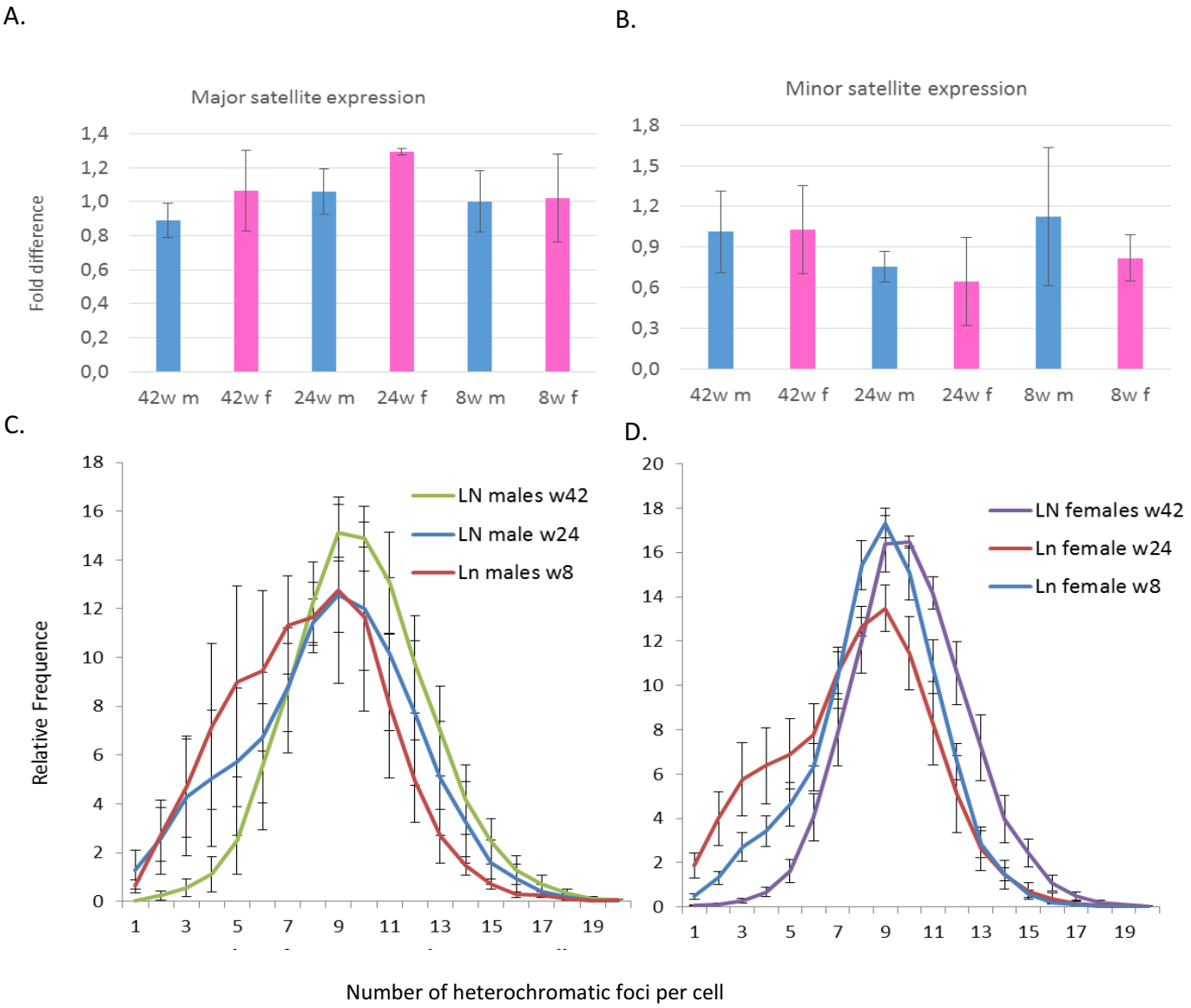


Figure 27 – Influence of age in chromatin structure and tandem repeat expression profile in lymph nodes.

Major satellite (A) and minor satellite (B) repeat expression in lymph nodes' lymphocytes of males and females at ages 8, 24 and 42 weeks. Heterochromatic foci distribution in males (C) and females (D) of the same age. Lymphocytes were stained with Draq5™ and analysed using ImageStreamX. Between 600 and 3000 cells were considered. Statistical significance was assessed via t-test (unpaired, 3 tails) and found no statistical significance. N=3 or 4 biological replicas; SEM=error bars. N=3-4 biological replicas; SEM=error bars.

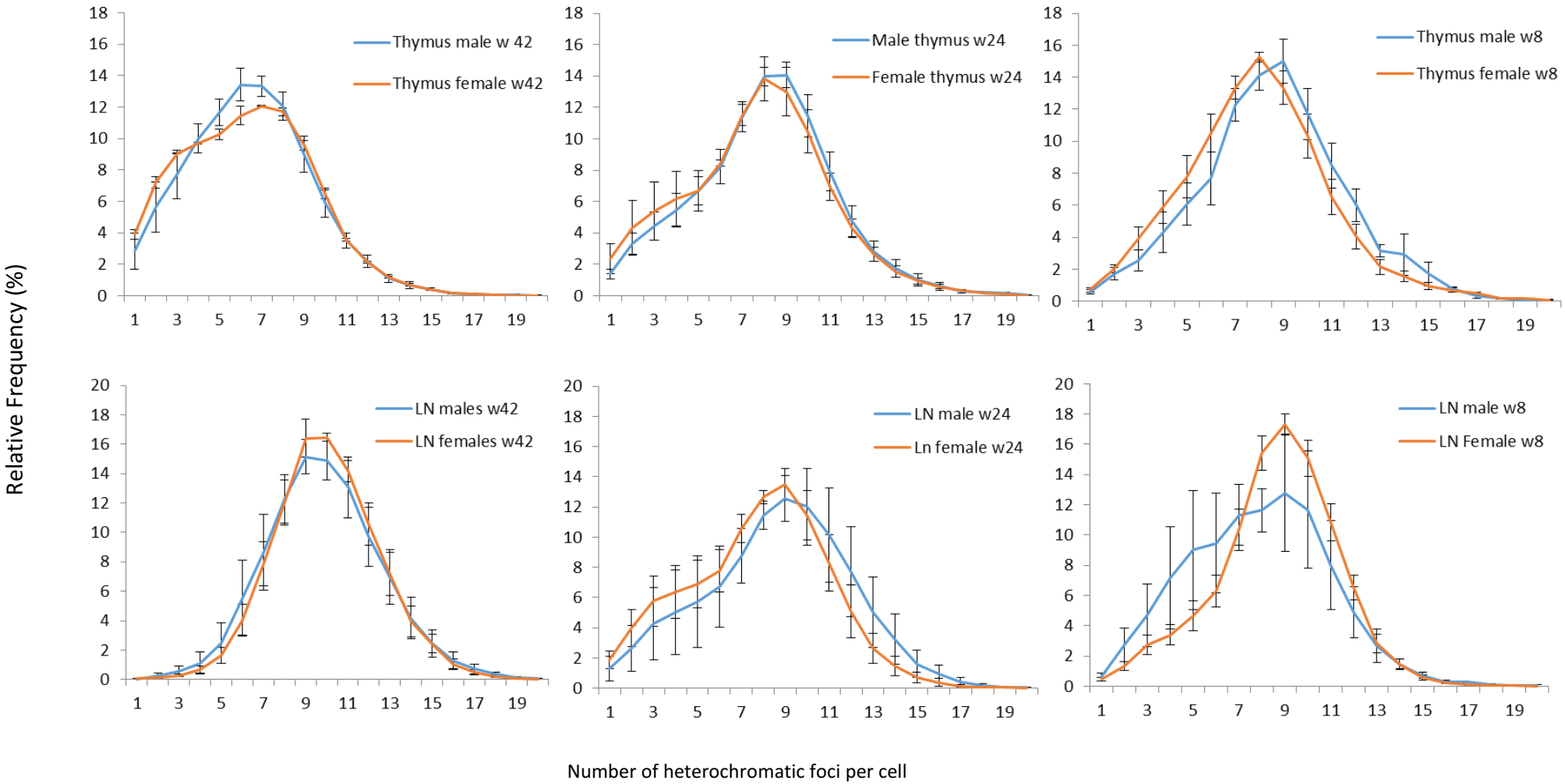


Figure 28- Heterochromatic foci distribution of males versus females, at 42, 24 and 8 weeks of age, in thymus and lymph nodes tissues (Draq5™).

Top three plots show distribution of heterochromatic foci on thymocytes per age group, whereas bottom three show, from the same mice, the lymph node tissue heterochromatic foci distribution. Cells were stained with Draq5™ and analysed using ImageStreamX. Between 300 and 8000 cells were considered. Frequence plots show no significant difference between genders in none of the age groups/tissues (sstatistical significance was assessed via t-test, unpaired, 3 tails). n=3 or 4 biological replicas; SEM=error bars.

Raquel Silva

4.1- Major satellites repeats influence gene expression of major satellite proximal genes in chromosome 9 and are testosterone sensitive

Analysis of the thymic transcriptome of FCG mice revealed that the presence of the Sry gene, the Y chromosome and testosterone (XY male), major and minor satellites were more repressed than in the other genotypes (XX males, XX and XY females). In contrast, presence of Sry on the XX male background resulted in upregulation of repeats compared to the XX female. Since XY and XX male mice have similar levels of testosterone, the only difference is their sex chromosome complement. This finding suggests a modulatory role of Sry, which is in agreement with our previous study (Wijchers et al. 2010) and allows us to speculate the existence of one (or more) factors regulating repeat expression on the Y chromosome. Other reports have proposed that Sry could act as (i) a repressor of anti-testis genes (McElreavey et al., 1993); (ii) as a structural factor responsible for bending DNA (Pontiggia et al., 1994); (iii) co-factor of gene silencing complexes (Oh, Li & Lau, 2005); (iv) as a splicing factor for mRNA (Ohe, Lalli & Sassone-Corsi, 2002); and more recently, (v) it has also been suggested that the circular transcript could act as a regulator of non-coding RNA (Hansen et al., 2013). However, to date none of these functions have been confirmed *in vivo*.

Sry is a known transcription factor whose main targets are in Sertoli cells (male genital ridges, testis) (Bullejos & Koopman, 2001) and in the brain: hypothalamus and cortex of human adult brain (Mayer et al., 1998; Clepet et al., 1993), and diencephalon, mesencephalon, and cortex in the mouse (Lahr et al., 1995). All cases described, Sry acts as a transactivator. However, its action in other cell types is much debated in the mouse, in part due to current methods that are poor at identifying expression of protein at lower levels. Reliable assessment of Sry expression is problematic at both the RNA and protein level. Thus, the Sry transcript is present in mouse cells as both linear and circular RNA transcripts (Capel et al., 1993), where the latter is thought to be sterile. With regard to detection of Sry protein, antibodies raised against Sry have been shown to cross-react with proteins from female cells (Festenstein lab, data not shown). An explanation for this may be its homology with 19 other members of the Sox family, in particular with autosomal Sox3.

In the FCG mice, an array of 13 copies of the Sry gene is present on chromosome 3, which is reported to cause a higher expression level compared to *wild type* XY male (Itoh et al., 2015). The thymus is not a known organ for Sry action, and although the RNA-seq picks up reads for Sry transcript, it is not known whether this transcript is linear or circular. Therefore, the observed Sry effect might come from either a direct or an indirect effect.

For this reason the potential effect of testosterone was investigated by removal of the male gonads. Testosterone was also a potential candidate for mediating the 'sry effect' as bioinformatics analysis revealed AR HRE in the major satellite DNA sequence. It was hypothesised that the effect on the pericentromere structure might influence the silencing of the adjacent centromeric repeats and cause minor satellite upregulation. The gonadectomy showed that XY males expressed these repeats at a higher level than the littermates with intact gonads, which strengthened the hypothesis that major satellite expression was testosterone sensitive. Analysis of the genomic position of genes that are differentially expressed as a result of gonadectomy identified a subset of these genes that co-localised with major satellite repeats. Interestingly, the comparison of XY males and XX females by the same method revealed that a subset of these sexually dimorphic DE genes were also testosterone-sensitive. These genes were not just adjacent to major satellites, they actually had parts of the major satellite monomer sequence within intronic and/or exonic regions. Together with the presence of an HRE for the androgen receptor, this suggests a direct testosterone effect in controlling the gene expression by means of satellite repeats. Interestingly, the AR complex is known to interact with known epigenetic modifiers (Shin & Janknecht, 2007), including some potential evidence that the Y encoded histone demethylase Kdm5d is one of these (A. Stratmann, PhD Thesis 2010).

Another very interesting feature of these major satellite proximal genes, is that some of the genes are very close to the pericentromeric region of chromosome 9 (Figure 9-A), suggesting a naturally occurring silencing phenomenon akin to position effect variegation. The effect of testosterone on these regions could thus be a mechanism to enhance heterochromatinisation. This contrasts with the human CD2 gene which is located near a heterochromatic region where its locus control region (LCR) is thought to prevent its silencing by avoiding heterochromatinisation (Festenstein et al., 1996). Another example of genes embedded in pericentromeric heterochromatin comes from studies in *Drosophila*, where the "light and rolled" genes apparently need heterochromatin in order to be expressed (Eberl, Duyf & Hilliker, 1993).

As the DE genes shown here to be close to major satellites have only recently been annotated little is currently known about their function which can only be hinted at by the presence of homology with known protein domains. Thus, the biological significance of the effects described await to be determined. It is tempting to speculate that, considering major satellites are more expressed in females, than males, and that expression of these repeats is associated with expression of the major satellite proximal genes, it suggests the interesting possibility that testosterone could silence genes necessary for female metabolism in the male background.

4.2- Heterochromatic foci number is not affected by testosterone

Analysis of heterochromatic foci can give insights into heterochromatin structure and stability as demonstrated by Zhu et al. (2011), where it was shown that ectopic expression of the heterochromatin repeat led to chromosome segregation anomalies and genome instability. In that report it was also shown that BRCA1 deficiency caused a reduction in the number of heterochromatic clusters in the nucleus, which was accompanied by upregulation of heterochromatic repeats, increased DNA damage response and apoptosis (Zhu et al., 2011). Although the authors reported the reduction in the number of centromere clusters and highlighted the correlation with increased expression of major satellite repeats the mechanism was not explored or discussed further. One possibility is that the increase in apoptosis, that they also demonstrated, is known to result in condensation of chromatin and therefore might have been responsible for the reduced number of clusters seen. In our study we were able to exclude apoptotic cells from the analysis using the live gate, which were very low in number in all cases, and we did not see a reduction in the number of centromeric clusters per cell. In addition, we observed here a trend towards an increasing number of heterochromatic foci per cell over time using DRAQ5 to label heterochromatin clusters. However, the number of heterochromatin clusters identified by HP1-GFP was unchanged. This implies that there might be a subset of heterochromatin ‘clusters’ identified by DRAQ5 which do not contain HP1. Moreover, the number of clusters identified by DRAQ5 was on average 1 spot more per cell than that identified by HP1-GFP. Such clusters might therefore represent regions of facultative heterochromatin. Further investigation of these regions might involve staining for H3K27me3 or polycomb components.

In the method developed here, living cells were used as opposed to fixed cells, and a much larger number of cells per group was considered than is practical using conventional microscopy. The applied gates also ensured a homogenous cell population in control versus gonadectomised groups, thus excluding apoptotic cells.

By using blood samples, it was also possible to assess any potential acute/chronic effect that may be caused by lack of testosterone.

The use of GFP-protein versus Draq5 staining had the advantage that no dye had to enter the cell and that it had previously been characterised as a *bone fide* marker for heterochromatic regions in T cells. The control group on the GFP cells analysis shows a far better overlap of distributions at different time points and smaller error bars than the controls on Draq5 analysis. There was some unexpected variation in the number of spots detected by Draq5 staining:, (i) the base-line measurement was different between control group and gonadectomised group, and (ii) the effect on sham mice on the last time measurement that equilibrated both groups to the same distribution and average number of spots per cell. At present there is no clear explanation for this having looked at possible technical features of the experiment including Draq5 concentration and cell number none of which seem to be responsible. Machine error is unlikely as the GFP-HP1 results do not show these effects.

Despite these concerns Draq5 staining at weeks 0, 2 and 6 did show a potentially interesting rearrangement of a steady increase of number heterochromatin foci per cell that would be interesting to follow up and the experiment will continue until 10 months at which time the mice will be culled and major satellite repeat expression assessed in thymocytes.

It has been previously shown that T-cells are testosterone sensitive *in vitro* decreasing lifespan in time and concentration dependent manner (McMurray et al., 2001), and gonadectomy in male rats was shown to increase the thymic cellularity (Leposavić et al., 1996). It will also be interesting to determine whether circulating lymphocytes show the same response to gonadectomy as tissue resident lymphocytes.

γH2AX assay could further assess the effect of gonadectomy on the DNA damage response which might be expected to have been activated given the previous demonstration that over expression of major satellite repeats induces DNA damage *in vitro* (Zhu et al., 2011). Additional measurements for chromosome segregation abnormalities (a feature of heterochromatin

disruption) could also be employed. Finally, gonadectomised mice could be rescued with testosterone and RNA FISH could allow visualization of major satellite expression location (pericentromeric regions).

4.3- Testosterone upregulated repeat expression *in vitro*

Given the interesting phenotype of the repressive effect of testosterone, it was attempted to reproduce this result in vitro. However, the result was surprisingly the opposite, and testosterone treatment of T-cells led to upregulation of major satellite repeats which did not show a classical dose response curve. Interestingly, the extent of upregulation of repeats was inversely proportional to the expression of Kdm5d, which may act to silence repeats by demethylating H3K4me2/3 suggesting a potential functional connection between the histone demethylase and testosterone. Interestingly, a preliminary experiment performed by M. Mauri indicated that knockdown of kdm5d in MEFs led to upregulation of major satellite repeats in the absence of testosterone. This major satellite repeat upregulation could also be the result of a partial agonist effect of testosterone, which has been described for spermatogenesis where it is inhibited by very high levels of testosterone.

In addition, the effect of culturing and activating T-cells cannot be excluded as being partially responsible for this paradoxical upregulatory effect on major satellite repeats. It would now be interesting to culture these cells for a longer time frame and compare repeat expression with non-activated cells directly. Intriguingly, in this regard,, male mouse embryonic fibroblast (MEFs) were also treated with testosterone and upregulated the major satellite repeats after 3 days of treatment and thereafter repressed their expression by day 7 compared with untreated (preliminary data not shown). Although preliminary, these results suggested that the *in vitro* models tested to study the phenomenon of repeat silencing seen in thymocytes in several important respects did not reflect the *in vivo* phenomenon emphasising the value of the *in vivo*/*ex vivo* approach.

4.4- The potential role of Kdm5d

As mentioned above Kdm5d is a potentially interesting Y chromosome candidate which might participate with SRY and/or testosterone in mediating repeat and gene silencing.

Previous studies showed that *Kdm5d* deletion occurred in 52% of human prostatic cancer cases analysed, while more recent studies indicate that low levels of expression of this demethylase were associated with worse prognosis (Perinchery et al., 2000), (Crea et al., 2012).

Although the mechanism is unknown, several independent clinical trials in prostate cancer make use of androgen deprivation followed by radiation therapy. With the information retrieved from the presented experiments, (even though these are based on a mouse model), it could be speculated that, the absence of testosterone and possible deletion of KDM5D could destabilize the heterochromatic structure and hence, make oncogenic cells more prone to the radiation. These studies could potentially help elucidate the mechanism behind the success of androgen deprivation and radiotherapy in prostatic cancer and could help optimize the treatment.

To address this hypothesis, as CRISPR has recently been shown to be capable of knocking out genes on the Y chromosome (Wang et al., 2013), Kdm5d could be knocked out in C57bl6 mice to investigate whether gonadectomy would destabilize chromatin structure and impair genome stability.

To further investigate whether Kdm5d plays a direct role in testosterone-mediated silencing in thymocytes, chromatin immunoprecipitation (ChIP) could be used to address whether Kdm5d and/or AR directly binds to satellite repeats at centromeric and pericentromeric heterochromatic regions and allow correlations with specific histone modifications such as H3K9me3 (a mark of constitutive heterochromatin) and H3K4me3 (the mark that Kdm5d demethylates). Interestingly, testosterone and/or Sry may participate directly in silencing the repeats as suggested by the presence of putative androgen receptor and the Sry binding sites found within the satellite repeats (M.Mauri, Master Thesis 2012).

4.5-Could the inactive X chromosome work as a heterochromatic sink in mammals?

The idea that the inactive X chromosome could serve as a heterochromatic sink comes from studies in *Drosophila*. In this genus, the Y chromosome is significantly larger than the X counterpart, and is largely heterochromatic. This has been proposed to result in sequestering of heterochromatinisation factors from other regions of the genome (Wijchers & Festenstein, 2011). In mammals, the inactive X chromosome is heterochromatinised early in embryonic development

and maintained throughout. One possibility is that, like the fly Y chromosome, the Barr body sequesters heterochromatic factors which results in less heterochromatinisation factors being available in the remaining chromatin regions, compared with cells with XO or XX background. This would imply that structures such as pericentromeres could be subjected to a reduced silencing effect thereby, leading to more expression of their repeats, as is the case in XX females and XY males of the FCG model. This was also observed in thymocytes from XO and XXY mice. Since the cross that originated these mice only produces XXY and XO mice. Mice with the same background XX and XY should be analysed to confirm this result. In summary, it seems that the more X chromosomes the more repeat expression.

4.6- SCC effects on heterochromatin gene expression might be due to genes escaping X-inactivation

Another interesting potential cause of the effect on SCC heterochromatin silencing could be due to the imbalance created by X inactivation escapees, including the epigenetic modifiers Kdm5c and Kdm6a. The former responsible for removing the active marks H3K4me2/3 and the latter, responsible for demethylation of the silencing marks H3K27me3, a facultative heterochromatin mark. Further *in vivo* studies by the use of knockout CRISPR approach could determine the effects on heterochromatic repeats and autosomal gene expression such as Kdm5c. Additionally, another potentially interesting gene to look at is the ATR X-linked gene, which is also an epigenetic regulator with evidence that it is biallelically expressed in trophoblast stem cells and has been implicated in regulating pericentromeric repeat heterochromatin in both humans and mice (Law et al., 2010), in the X chromosome inactivation process (Sarma et al., 2014) and spermatogenesis via AR receptor in sertoli cells (Bagheri-Fam et al., 2010).

4.7- Age and not gender affects repeat expression and heterochromatin structure in thymocytes.

Given that the two transcriptomic experiments on the FCG were done on mice of differing ages and revealed a different set of sex-chromosome sensitive complement genes it was decided to investigate the effect of age and gender on major satellite repeat expression and heterochromatic foci distribution in the nucleus, both in the thymus and lymph nodes tissues obtained from wild

type mice. The wild type background chosen was C57Bl6 (bred at Charles River), given that the FCG mice were also C57Bl6 (although bred at the Jackson's lab).

Although, in general there was more major satellite expression in female thymus compared with male (8 and 24 weeks) this was not statistically significant whereas expression in mature lymphocytes (lymph node) showed no difference. However, there was a reduction in major satellite expression at 42 weeks which was accompanied by a reduction in the number of heterochromatic foci per cell in both male and female thymocytes. This is consistent with a model in which as the thymus involutes with age the chromatin condenses and the number of heterochromatin clusters decreases and this correlates with reduced heterochromatin expression.

The fact that major satellite expression in thymus was different from that in lymph nodes, implies a tissue specific difference. An inverse correlation was also found with increasing age between the number of heterochromatic foci per cell in thymus which decreased in number, versus lymph node where there was an increase in number. This could be related to the known age-related thymic involution and possibly comes from the fact that the thymus has mainly immature T-cells, that are dividing and maturing into DP thymocytes, whereas in lymph nodes, ~70% of cells are mature T-cells and ~30% are B-cells (Supplementary 6.2), which are in a quiescent state. Since major satellite repeat expression is dependent on cell cycling and is repressed in quiescent cells (Chen et al., 2008) this could have contributed to the difference observed between tissues.

In female thymocytes, the repeat expression is highest at 24 weeks. This could be due to the hormone regulation, since by 42 weeks the mice are premenopausal (Beagley & Gockel, 2003). This would be consistent with the finding that removal of ovaries from XX mice also led to downregulation of the heterochromatic repeats (Fig 11).

Surprisingly, the qRT-PCR repeat expression results obtained from 42 week old mice presented here did not entirely agree with the RNA seq expression analysis previously performed on FCG mice of the same age. In the FCG experiment the males (XY) expressed major satellites at a lower level than females (XX) whereas in the experiments shown here the wild type males expressed more major satellites than the wild type females. At this age the females are peri-menopausal and one possibility is that the C57Black6/Charles River mice which were bred in a different animal house underwent their menopause earlier than the FCG C57Black6/Jax mice, in this case the effect of reduced female hormones would be to repress heterochromatin expression to greater extent

than the males (Figs 12 and 26). This could be confirmed by assessment of hormone levels. Because males and females are in fact healthy cells, with similar cell cycle population, one would not expect a very big difference in the stability of the heterochromatic structure. In this thesis it was shown that removal of hormones can have effects on the stability silencing of major satellites and that these effects are sexually dimorphic. On the other hand, epigenetic alterations have been shown as hallmarks of aging, including enlarged nuclei, loss of heterochromatin factors, such as HP1 proteins and loss of H3K9me3 (Zhang et al., 2015; Faial, 2015), that could thus destabilize the structure and contribute to the changes in heterochromatic foci observed which were apparently not sexually dimorphic .

4.8- Conclusions and future directions

In summary, described here is the development of a novel methodology to track the number of heterochromatin clusters in large numbers of living cells. Taking advantage of transgenic mice which express a GFP labelled version of HP1β in T cells it was shown that the labelled heterochromatic foci were stable in number for 12 weeks post gonadectomy. However, preliminary data using Draq5 in circulating lymphocytes identified heterochromatin foci and showed a trend towards an increased number of foci post-gonadectomy. These results imply that HP1 deficient heterochromatin regions might be sensitive to lack of testosterone. However, this hypothesis needs to be addressed with further experiments to ensure reliability of the technique and investigate the nature of these potentially facultative heterochromatin foci in more detail. Draq5 labelling also showed age-related and tissue-specific changes in the number of heterochromatin foci whereby with increasing age the involuting thymus decreased the number of foci whereas the quiescent mature T cells had a trend towards an increase in their number. This might be a reflection of the difference in cell cycle and maturity of the cells.

Transcriptomic analysis revealed regions of major satellite repeats on chromosome 2 and 9 within which were a subset of 14 genes whose expression is sexually dimorphic and repressed by testosterone. Interestingly, these genes have only recently been annotated and several have the repeat sequence embedded in both exonic and intronic regions. Analysis of their expression between the different genotypes of the FCG model revealed that they precisely mirrored the effect seen on the major satellite repeat. Therefore these genes appear to be the first example of genes directly influenced by their proximity to constitutive heterochromatin in a manner that

resembles the archetypal epigenetic phenomenon of position effect variegation. Moreover, it was shown here that testosterone might silence such genes that are regulated by heterochromatin. Notably many of these genes were expressed at a high level and the effects on transcription showed a fold change up to 3.78 ($P<0.0001$). It remains to be seen what the precise function of the genes is however, bioinformatic analysis of the predicted protein structure showed that several of them had homology with metabolic enzymes a category which is known to be sexually dimorphic.

5-REFERENCES

Anders, S. (2010) Analysing RNA-Seq data with the DESeq package. *Mol Biol.* 43 (4), 1-17.

Ando, S., Yang, H., Nozaki, N., Okazaki, T. & Yoda, K. (2002) CENP-A, -B, and -C chromatin complex that contains the I-type alpha-satellite array constitutes the prekinetochore in HeLa cells. *Molecular and Cellular Biology.* 22 (7), 2229-2241.

Arnold, A. P. (2014) Conceptual frameworks and mouse models for studying sex differences in physiology and disease: Why compensation changes the game. *Experimental Neurology.*

Arnold, A. P. (2012) The end of gonad-centric sex determination in mammals. *Trends in Genetics.* 28 (2), 55-61.

Arnold, A. P. (2009) The organizational–activational hypothesis as the foundation for a unified theory of sexual differentiation of all mammalian tissues. *Hormones and Behavior.* 55 (5), 570-578.

Bagheri-Fam, S., Argentaro, A., Svingen, T., Combes, A., Koopman, P. & Harley, V. (2010) ATRX is a sertoli cell survival factor and regulator of spermatogenesis via interaction with androgen receptor. *Endocrine Reviews.* , The Endocrine Society.

Bannister, A. J. & Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Research.* 21 (3), 381-395.

Barbot, W., Dupressoir, A., Lazar, V. & Heidmann, T. (2002) Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. *Nucleic Acids Research.* 30 (11), 2365-2373.

Barr, M. L. & Moore, K. L. (1957) Chromosomes, sex chromatin, and cancer. *Proceedings.Canadian Cancer Conference.* 23-16.

Beagley, K. W. & Gockel, C. M. (2003) Regulation of innate and adaptive immunity by the female sex hormones oestradiol and progesterone. *FEMS Immunology & Medical Microbiology.* 38 (1), 13-22.

Brown, G. R. & Spencer, K. A. (2012) Steroid hormones, stress and the adolescent brain: a comparative perspective. *Neuroscience.*

Bullejos, M. & Koopman, P. (2001) Spatially dynamic expression of Sry in mouse genital ridges. *Developmental Dynamics.* 221 (2), 201-205.

Burgoyne, P., Thornhill, A., Boudrean, S. K., Darling, S., Bishop, C., Evans, E., Capel, B. & Mittwoch, U. (1995) The Genetic Basis of XX-XY Differences Present before Gonadal Sex Differentiation in

the Mouse [and Discussion]. *Philosophical Transactions of the Royal Society of London.Series B: Biological Sciences.* 350 (1333), 253-261.

Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. & Lovell-Badge, R. (1993) Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell.* 73 (5), 1019-1030.

Carone, D. M. & Lawrence, J. B. (2012) Heterochromatin instability in cancer: From the Barr body to satellites and the nuclear periphery. *Seminars in Cancer Biology.* , Elsevier.

Casa, V. & Gabellini, D. (2012) A repetitive elements perspective in Polycomb epigenetics. *Frontiers in Genetics.* 3.

Castanotto, D., Tommasi, S., Li, M., Li, H., Yanow, S., Pfeifer, G. P. & Rossi, J. J. (2005) Short hairpin RNA-directed cytosine (CpG) methylation of the RASSF1A gene promoter in HeLa cells. *Molecular Therapy.* 12 (1), 179-183.

Cedar, H. & Bergman, Y. (2012) Programming of DNA methylation patterns. *Annual Review of Biochemistry.* 8197-117.

Chen, E. S., Zhang, K., Nicolas, E., Cam, H. P., Zofall, M. & Grewal, S. I. (2008) Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature.* 451 (7179), 734-737.

Chen, X., McClusky, R., Chen, J., Beaven, S. W., Tontonoz, P., Arnold, A. P. & Reue, K. (2012) The number of X chromosomes causes sex differences in adiposity in mice. *PLoS Genetics.* 8 (5), e1002709.

Clepet, C., Schafer, A. J., Sinclair, A. H., Palmer, M. S., Lovell-Badge, R. & Goodfellow, P. N. (1993) The human SRY transcript. *Human Molecular Genetics.* 2 (12), 2007-2012.

Crea, F., Sun, L., Mai, A., Chiang, Y. T., Farrar, W. L., Danesi, R. & Helgason, C. D. (2012) The emerging role of histone lysine demethylases in prostate cancer. *Molecular Cancer.* 11 (1), 1-10.

de Koning, A. J., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genetics.* 7 (12), e1002384.

De Vries, G. J., Rissman, E. F., Simerly, R. B., Yang, L. Y., Scordalakes, E. M., Auger, C. J., Swain, A., Lovell-Badge, R., Burgoyne, P. S. & Arnold, A. P. (2002) A model system for study of sex chromosome effects on sexually dimorphic neural and behavioral traits. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience.* 22 (20), 9005-9014.

Dillon, N. (2004) Heterochromatin structure and function. *Biology of the Cell.* 96 (8), 631-637.

Eberl, D. F., Duyf, B. J. & Hilliker, A. J. (1993) The role of heterochromatin in the expression of a heterochromatic gene, the rolled locus of Drosophila melanogaster. *Genetics.* 134 (1), 277-292.

Ekram, M. B., Kang, K., Kim, H. & Kim, J. (2012) Retrotransposons as a major source of epigenetic variations in the mammalian genome. *Epigenetics.* 7 (4), 370-382.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. & Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature.* 411 (6836), 494-498.

Ellegren, H. & Parsch, J. (2007) The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics.* 8 (9), 689-698.

Faial, T. (2015) Heterochromatin and human aging. *Nature Genetics.* 47 (6), 568-568.

Festenstein, R., Pagakis, S. N., Hiragami, K., Lyon, D., Verreault, A., Sekkali, B. & Kioussis, D. (2003) Modulation of heterochromatin protein 1 dynamics in primary mammalian cells. *Science.* 299 (5607), 719-721.

Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M. & Kioussis, D. (1996) Locus control region function and heterochromatin-induced position effect variegation. *Science.* 271 (5252), 1123-1125.

Frescas, D., Guardavaccaro, D., Kuchay, S. M., Kato, H., Poleshko, A., Basrur, V., Elenitoba-Johnson, K. S., Katz, R. A. & Pagano, M. (2008) KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. *Cell Cycle.* 7 (22), 3539-3547.

Gatewood, J. D., Wills, A., Shetty, S., Xu, J., Arnold, A. P., Burgoyne, P. S. & Rissman, E. F. (2006) Sex chromosome complement and gonadal sex influence aggressive and parental behaviors in mice. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience.* 26 (8), 2335-2342.

Gentilini, D., Mari, D., Castaldi, D., Remondini, D., Ogliari, G., Ostan, R., Bucci, L., Sirchia, S. M., Tabano, S. & Cavagnini, F. (2013) Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring. *Age.* 35 (5), 1961-1973.

Grewal, S. I. & Jia, S. (2007) Heterochromatin revisited. *Nature Reviews Genetics.* 8 (1), 35-46.

Grunstein, M. (1997) Histone acetylation in chromatin structure and transcription. *Nature.* 389 (6649), 349-352.

Guerriero, G. (2009) Vertebrate sex steroid receptors: evolution, ligands, and neurodistribution. *Annals of the New York Academy of Sciences.* 1163 (1), 154-168.

Hall, L. E., Mitchell, S. E. & O'Neill, R. J. (2012) Pericentric and centromeric transcription: a perfect balance required. *Chromosome Research.* 20 (5), 535-546.

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K. & Kjems, J. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature.* 495 (7441), 384-388.

Herrmann, J. L., Abarbanell, A. M., Weil, B. R., Manukyan, M. C., Poynter, J. A., Wang, Y., Coffey, A. C. & Meldrum, D. R. (2010) Gender dimorphisms in progenitor and stem cell function in cardiovascular disease. *Journal of Cardiovascular Translational Research.* 3 (2), 103-113.

Hsieh, C., Lin, C., Liu, H., Chang, Y., Shih, C., Zhong, C. Z., Lee, S. & Tan, B. C. (2011) WDHD1 modulates the post-transcriptional step of the centromeric silencing pathway. *Nucleic Acids Research.* 39 (10), 4048-4062.

Itoh, Y., Mackie, R., Kampf, K., Domadia, S., Brown, J. D., O'Neill, R. & Arnold, A. P. (2015) Four Core Genotypes mouse model: localization of the Sry transgene and bioassay for testicular hormone levels. *BMC Research Notes.* 8 (1), 69.

Jazaeri, A. A., Yee, C. J., Sotiriou, C., Brantley, K. R., Boyd, J. & Liu, E. T. (2002) Gene expression profiles of BRCA1-linked, BRCA2-linked, and sporadic ovarian cancers. *Journal of the National Cancer Institute.* 94 (13), 990-1000.

Jenuwein, T. & Allis, C. D. (2001) Translating the histone code. *Science.* 293 (5532), 1074-1080.

Ji, H., Zheng, W., Wu, X., Liu, J., Ecelbarger, C. M., Watkins, R., Arnold, A. P. & Sandberg, K. (2010) Sex Chromosome Effects Unmasked in Angiotensin II–Induced Hypertension. *Hypertension.* 55 (5), 1275-1282.

Jolly, C., Metz, A., Govin, J., Vigneron, M., Turner, B. M., Khochbin, S. & Vourc'h, C. (2004) Stress-induced transcription of satellite III repeats. *The Journal of Cell Biology.* 164 (1), 25-33.

Jousilahti, P., Vartiainen, E., Tuomilehto, J. & Puska, P. (1999) Sex, age, cardiovascular risk factors, and coronary heart disease: a prospective follow-up study of 14 786 middle-aged men and women in Finland. *Circulation.* 99 (9), 1165-1172.

Kang, N., Ha, H., Yun, S., Yu, Y. H. & Chang, Y. (2011) Diversity-driven chemical probe development for biomolecules: beyond hypothesis-driven approach. *Chemical Society Reviews.* 40 (7), 3613-3626.

Kisielow, P., Teh, H. S., Blüthmann, H. & von Boehmer, H. (1988) Positive selection of antigen-specific T cells in thymus by restricting MHC molecules. *Nature.* 335 (6192), 730-733.

Kosztin, D., Bishop, T. C. & Schulten, K. (1997) Binding of the estrogen receptor to DNA. The role of waters. *Biophysical Journal.* 73 (2), 557-570.

Kouzarides, T. (2007) Chromatin modifications and their function. *Cell.* 128 (4), 693-705.

Kuroki, S., Matoba, S., Akiyoshi, M., Matsumura, Y., Miyachi, H., Mise, N., Abe, K., Ogura, A., Wilhelm, D., Koopman, P., Nozaki, M., Kanai, Y., Shinkai, Y. & Tachibana, M. (2013) Epigenetic regulation of mouse sex determination by the histone demethylase Jmjd1a. *Science (New York, N.Y.).* 341 (6150), 1106-1109.

Kwon, S. H. & Workman, J. L. (2011) The changing faces of HP1: From heterochromatin formation and gene silencing to euchromatic gene expression. *Bioessays.* 33 (4), 280-289.

Lahr, G., Maxson, S. C., Mayer, A., Just, W., Pilgrim, C. & Reisert, I. (1995) Transcription of the Y chromosomal gene, Sry, in adult mouse brain. *Molecular Brain Research.* 33 (1), 179-182.

Lange, U. C., Siebert, S., Wossidlo, M., Weiss, T., Ziegler-Birling, C., Walter, J., Torres-Padilla, M., Daujat, S. & Schneider, R. (2013) Dissecting the role of H3K64me3 in mouse pericentromeric heterochromatin. *Nature Communications.* 4.

Law, M. J., Lower, K. M., Voon, H. P., Hughes, J. R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M. & Morris, A. (2010) ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell.* 143 (3), 367-378.

Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., Lohr, J. G., Harris, C. C., Ding, L. & Wilson, R. K. (2012) Landscape of somatic retrotransposition in human cancers. *Science.* 337 (6097), 967-971.

Leposavić, G., Karapetrović, B., Obradović, S., Danković, B. V. & Kosec, D. (1996) Differential effects of gonadectomy on the thymocyte phenotypic profile in male and female rats. *Pharmacology Biochemistry and Behavior.* 54 (1), 269-276.

Lopez-Contreras, A. J. & Fernandez-Capetillo, O. (2012) Signalling DNA Damage.

Lovell-Badge, R. & Robertson, E. (1990) XY female mice resulting from a heritable mutation in the primary testis-determining gene, Tdy. *Development.* 109 (3), 635-646.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997a) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature.* 389 (6648), 251-260.

Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. & Richmond, T. J. (1997b) Characterization of nucleosome core particles containing histone proteins made in bacteria. *Journal of Molecular Biology.* 272 (3), 301-311.

Mahadevaiah, S. K., Odorisio, T., Elliott, D. J., Rattigan, Á, Szot, M., Laval, S. H., Washburn, L. L., McCarrey, J. R., Cattanach, B. M. & Lovell-Badge, R. (1998) Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities. *Human Molecular Genetics.* 7 (4), 715-727.

Maison, C., Bailly, D., Peters, A. H., Quivy, J., Roche, D., Taddei, A., Lachner, M., Jenuwein, T. & Almouzni, G. (2002) Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nature Genetics.* 30 (3), 329-334.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. (1997) Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes & Development.* 11 (2), 156-166.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N.,

Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C. & Bryant, S. H. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Research.* 43 (Database issue), D222-6.

Martens, J. H., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P. & Jenuwein, T. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *The EMBO Journal.* 24 (4), 800-812.

Matzke, M., Mette, M. & Matzke, A. (2000) Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. In: Anonymous *Plant Gene Silencing.* , Springer. pp. 281-295.

Mayer, A., Lahr, G., Swaab, D. F., Pilgrim, C. & Reisert, I. (1998) The Y-chromosomal genes SRY and ZFY are transcribed in adult human brain. *Neurogenetics.* 1 (4), 281-288.

McElreavey, K., Vilain, E., Abbas, N., Herskowitz, I. & Fellous, M. (1993) A regulatory cascade hypothesis for mammalian sex determination: SRY represses a negative regulator of male development. *Proceedings of the National Academy of Sciences of the United States of America.* 90 (8), 3368-3372.

McMurray, R. W. (2001) Estrogen, prolactin, and autoimmunity: actions and interactions. *International Immunopharmacology.* 1 (6), 995-1008.

McMurray, R. W., Suwannaroj, S., Ndebele, K. & Jenkins, J. K. (2001) Differential effects of sex steroids on T and B cells: modulation of cell cycle phase distribution, apoptosis and bcl-2 protein levels. *Pathobiology : Journal of Immunopathology, Molecular and Cellular Biology.* 69 (1), 44-58.

McPhie-Lalmansingh, A. A., Tejada, L. D., Weaver, J. L. & Rissman, E. F. (2008) Sex chromosome complement affects social interactions in mice. *Hormones and Behavior.* 54 (4), 565-570.

Mittelstrass, K., Ried, J. S., Yu, Z., Krumsiek, J., Gieger, C., Prehn, C., Roemisch-Margl, W., Polonikov, A., Peters, A. & Theis, F. J. (2011) Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genetics.* 7 (8), e1002215.

Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. (2004) Small interfering RNA-induced transcriptional gene silencing in human cells. *Science.* 305 (5688), 1289-1292.

Muller, H. (1930) Types of visible variations induced by X-rays inDrosophila. *Journal of Genetics.* 22 (3), 299-334.

Munger, S. C. & Capel, B. (2012) Sex and the circuitry: progress toward a systems-level understanding of vertebrate sex determination. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine.* 4 (4), 401-412.

Muotri, A. R., Chu, V. T., Marchetto, M. C., Deng, W., Moran, J. V. & Gage, F. H. (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature.* 435 (7044), 903-910.

Narayan, A., Ji, W., Zhang, X., Marrogi, A., Graff, J. R., Baylin, S. B. & Ehrlich, M. (1998) Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *International Journal of Cancer.* 77 (6), 833-838.

Neguembor, M. V. & Gabellini, D. (2010) In junk we trust: repetitive DNA, epigenetics and facioscapulohumeral muscular dystrophy. *Epigenomics.* 2 (2), 271-287.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D. & Grewal, S. I. (2004) RITS acts in cis to promote RNA interference–mediated transcriptional and post-transcriptional silencing. *Nature Genetics.* 36 (11), 1174-1180.

Oh, H. J., Li, Y. & Lau, Y. F. (2005) Sry associates with the heterochromatin protein 1 complex by interacting with a KRAB domain protein. *Biology of Reproduction.* 72 (2), 407-415.

Ohe, K., Lalli, E. & Sassone-Corsi, P. (2002) A direct role of SRY and SOX proteins in pre-mRNA splicing. *Proceedings of the National Academy of Sciences.* 99 (3), 1146-1151.

Oliveros, J. (2008) *VENNY.an Interactive Tool for Comparing Lists with Venn Diagrams.2007.*

Pasquali, R. (2006) Obesity and androgens: facts and perspectives. *Fertility and Sterility.* 85 (5), 1319-1340.

Pedersen, M. T. & Helin, K. (2010) Histone demethylases in development and disease. *Trends in Cell Biology.* 20 (11), 662-671.

Perinchery, G., SASAKI, M., ANGAN, A., KUMAR, V., CARROLL, P. & DAHIYA, R. (2000) Deletion of Y-chromosome specific genes in human prostate cancer. *The Journal of Urology.* 163 (4), 1339-1342.

Piferrer, F. (2013) Epigenetics of sex determination and gonadogenesis. *Developmental Dynamics.*

Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R. & Bianchi, M. E. (1994) Sex-reversing mutations affect the architecture of SRY-DNA complexes. *The EMBO Journal.* 13 (24), 6115-6124.

Pusarla, R. & Bhargava, P. (2005) Histones in functional diversification. *FEBS Journal.* 272 (20), 5149-5168.

Richardson, A. L., Wang, Z. C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J. D., Livingston, D. M. & Ganesan, S. (2006) X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell.* 9 (2), 121-132.

Roy, A. K. & Chatterjee, B. (1983) Sexual dimorphism in the liver. *Annual Review of Physiology.* 45 (1), 37-50.

Sakiani, S., Olsen, N. J. & Kovacs, W. J. (2012) Gonadal steroids and humoral immunity. *Nature Reviews Endocrinology.*

Sarachana, T. & Hu, V. W. (2013) Genome-wide identification of transcriptional targets of RORA reveals direct regulation of multiple genes associated with autism spectrum disorder. *Molecular Autism.* 4 (1), 14.

Sarma, K., Cifuentes-Rojas, C., Ergun, A., del Rosario, A., Jeon, Y., White, F., Sadreyev, R. & Lee, J. T. (2014) ATRX Directs Binding of PRC2 to Xist RNA and Polycomb Targets. *Cell.* 159 (4), 869-883.

Sasidhar, M. V., Itoh, N., Gold, S. M., Lawson, G. W. & Voskuhl, R. R. (2012) The XX sex chromosome complement in mice is associated with increased spontaneous lupus compared with XY. *Annals of the Rheumatic Diseases.* 71 (8), 1418-1422.

Schulz, J. B., Boesch, S., Bürk, K., Dürr, A., Giunti, P., Mariotti, C., Pousset, F., Schöls, L., Vankan, P. & Pandolfo, M. (2009) Diagnosis and treatment of Friedreich ataxia: a European perspective. *Nature Reviews Neurology.* 5 (4), 222-234.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A. & Shi, Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell.* 119 (7), 941-953.

Shin, S. & Janknecht, R. (2007) Activation of androgen receptor by histone demethylases JMJD2A and JMJD2D. *Biochemical and Biophysical Research Communications.* 359 (3), 742-746.

Smith-Bouvier, D. L., Divekar, A. A., Sasidhar, M., Du, S., Tiwari-Woodruff, S. K., King, J. K., Arnold, A. P., Singh, R. R. & Voskuhl, R. R. (2008) A role for sex chromosome complement in the female bias in autoimmune disease. *The Journal of Experimental Medicine.* 205 (5), 1099-1108.

Solyom, S. & Kazazian Jr, H. H. (2012) Mobile elements in the human genome: implications for disease. *Genome Medicine.* 4 (2), 1-8.

Strahl, B. D. & Allis, C. D. (2000) The language of covalent histone modifications. *Nature.* 403 (6765), 41-45.

Taft, R. J., Pheasant, M. & Mattick, J. S. (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays.* 29 (3), 288-299.

Thomas, P., Pang, Y., Dong, J., Groenen, P., Kelder, J., De Vlieg, J., Zhu, Y. & Tubbs, C. (2007) Steroid and G protein binding characteristics of the seatrout and human progestin membrane receptor α subtypes and their evolutionary origins. *Endocrinology.* 148 (2), 705-718.

Turner, B. M. (2009) Epigenetic responses to environmental change and their evolutionary implications. *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences.* 364 (1534), 3403-3418.

van Nas, A., GuhaThakurta, D., Wang, S. S., Yehya, N., Horvath, S., Zhang, B., Ingram-Drake, L., Chaudhuri, G., Schadt, E. E. & Drake, T. A. (2009) Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks. *Endocrinology.* 150 (3), 1235-1249.

Volpe, T. & Martienssen, R. A. (2011) RNA interference and heterochromatin assembly. *Cold Spring Harbor Perspectives in Biology.* 3 (9), .

Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F. & Jaenisch, R. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell.* 153 (4), 910-918.

Wijchers, P. J. & Festenstein, R. J. (2011) Epigenetic regulation of autosomal gene expression by sex chromosomes. *Trends in Genetics.* 27 (4), 132-140.

Yamamoto, K. & Sonoda, M. (2003) Self-interaction of heterochromatin protein 1 is required for direct binding to histone methyltransferase, SUV39H1. *Biochemical and Biophysical Research Communications.* 301 (2), 287-292.

Yandim, C., Natisvili, T. & Festenstein, R. (2013) Gene regulation and epigenetics in Friedreich's ataxia. *Journal of Neurochemistry.* 126 (s1), 21-42.

Yang, X., Schadt, E. E., Wang, S., Wang, H., Arnold, A. P., Ingram-Drake, L., Drake, T. A. & Lusis, A. J. (2006) Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Research.* 16 (8), 995-1004.

Zaidi, S. K., Young, D. W., Montecino, M., Lian, J. B., Stein, J. L., van Wijnen, A. J. & Stein, G. S. (2010) Architectural epigenetics: mitotic retention of mammalian transcriptional regulatory information. *Molecular and Cellular Biology.* 30 (20), 4758-4766.

Zeng, W., Ball Jr, A. R. & Yokomori, K. (2010) HP1: heterochromatin binding proteins working the genome. *Epigenetics.* 5 (4), 287-292.

Zhang, W., Li, J., Suzuki, K., Qu, J., Wang, P., Zhou, J., Liu, X., Ren, R., Xu, X., Ocampo, A., Yuan, T., Yang, J., Li, Y., Shi, L., Guan, D., Pan, H., Duan, S., Ding, Z., Li, M., Yi, F., Bai, R., Wang, Y., Chen, C., Yang, F., Li, X., Wang, Z., Aizawa, E., Goebl, A., Soligalla, R. D., Reddy, P., Esteban, C. R., Tang, F., Liu, G. H. & Belmonte, J. C. (2015) Aging stem cells. A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science (New York, N.Y.).* 348 (6239), 1160-1163.

Zhao, G. J., Koopmans, R. A., Li, D. K., Bedell, L. & Paty, D. W. (2000) Effect of interferon beta-1b in MS: assessment of annual accumulation of PD/T2 activity on MRI. UBC MS/MRI Analysis Group and the MS Study Group. *Neurology.* 54 (1), 200-206.

Zhu, Q., Pao, G. M., Huynh, A. M., Suh, H., Tonnu, N., Nederlof, P. M., Gage, F. H. & Verma, I. M. (2011) BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature.* 477 (7363), 179-184.

Zvetkova, I., Apedaile, A., Ramsahoye, B., Mermoud, J. E., Crompton, L. A., John, R., Feil, R. & Brockdorff, N. (2005) Global hypomethylation of the genome in XX embryonic stem cells. *Nature Genetics.* 37 (11), 1274-1279.

6-SUPPLEMENTARY

6.1-DE expressed genes in XY and XX Females (FCG)

Note:

a) in yellow – common DE genes between the four comparisons (in red – DE gene with predicted homology for protein domain with progesterone binding region;

b) in blue DE genes in both XY male intact vs XY male GDX and XX GDX female vs XY GDX male;

c) in green – DE gene specific from XY intact male vs XY gonadectomised male.

XY intact male vs XY gonadectomised male: DE expressed genes with zero distance to major satellite

seqnames,"start","end","width","strand","name","biotype","Feature","DistToSat"
chr2,98502394,98504240,1847,"+","ENSMUSG00000075015","protein_coding","Gene",0
chr9,3023547,3025218,1672,"+","ENSMUSG00000074562","protein_coding","Gene",0
chr9,3017408,3019022,1615,"+","ENSMUSG00000074563","protein_coding","Gene",0
chr9,3015654,3017210,1557,"+","ENSMUSG00000074564","protein_coding","Gene",0
chr9,3020155,3021593,1439,"+","ENSMUSG00000079719","protein_coding","Gene",0
chr9,3000922,3002330,1409,"+","ENSMUSG00000091028","protein_coding","Gene",0
chr9,3034599,3035805,1207,"+","ENSMUSG00000091057","protein_coding","Gene",0
chr9,3037111,3038316,1206,"+","ENSMUSG00000074558","protein_coding","Gene",0
chr9,3013170,3014344,1175,"+","ENSMUSG00000074565","protein_coding","Gene",0
chr9,3025417,3026387,971,"+","ENSMUSG00000074561","protein_coding","Gene",0
chr2,98506704,98507458,755,"-","ENSMUSG00000075014","protein_coding","Gene",0
chr9,3004457,3005194,738,"+","ENSMUSG00000090650","protein_coding","Gene",0
chr9,3032085,3032822,738,"+","ENSMUSG00000061971","protein_coding","Gene",0
chr9,3030337,3031071,735,"+","ENSMUSG00000091599","protein_coding","Gene",0

XY intact male vs XY gonadectomised male: DE genes more than1MB distant from repeats

seqnames,"start","end","width","strand","name","biotype","Feature","DistToSat"
chr2,125126330,125333729,207400,"-","ENSMUSG00000027204","protein_coding","Gene",26618840
chr5,51845488,51958964,113477,"-","ENSMUSG00000029167","protein_coding","Gene",52626541
chr5,17287508,17394777,107270,"-","ENSMUSG00000002944","protein_coding","Gene",87190728
chr2,58119864,58210306,90443,"-","ENSMUSG00000026834","protein_coding","Gene",40292084
chr8,47556395,47621405,65011,"+","ENSMUSG00000018796","protein_coding","Gene",23350117
chrX,137074067,137135198,61132,"-","ENSMUSG00000031431","protein_coding","Gene",64156185
chr11,104411977,104457067,45091,"+","ENSMUSG00000061086","protein_coding","Gene",101407224
chr6,4454814,4491543,36730,"+","ENSMUSG00000029661","protein_coding","Gene",38802216
chr2,126378143,126413979,35837,"+","ENSMUSG00000027359","protein_coding","Gene",27870653

chr6,17256335,17291452,35118,"+","ENSMUSG00000007655","protein_coding","Gene",26002307
chr8,95689967,95721737,31771,"-","ENSMUSG00000056973","protein_coding","Gene",71483689
chr8,71404390,71431347,26958,"+","ENSMUSG00000015568","protein_coding","Gene",47198112
chr11,54971972,54998579,26608,"-","ENSMUSG00000020264","protein_coding","Gene",51967219
chr6,5118090,5143946,25857,"-","ENSMUSG00000002588","protein_coding","Gene",38149813
chr18,67503218,67527448,24231,"+","ENSMUSG00000024526","protein_coding","Gene",11403917
chr11,67051314,67073948,22635,"+","ENSMUSG00000057003","protein_coding","Gene",64046561
chrX,130683582,130696615,13034,"-","ENSMUSG00000031258","protein_coding","Gene",57765700
chr6,5433351,5446309,12959,"-","ENSMUSG00000019577","protein_coding","Gene",37847450
chr16,23146609,23158101,11493,"+","ENSMUSG00000022878","protein_coding","Gene",12171232
chr3,14863538,14872351,8814,"+","ENSMUSG00000027559","protein_coding","Gene",41507004
chr8,85814247,85822355,8109,"+","ENSMUSG00000031710","protein_coding","Gene",61607969
chr11,54715955,54723879,7925,"+","ENSMUSG00000018339","protein_coding","Gene",51711202
chr19,36186455,36194334,7880,"-","ENSMUSG00000024803","protein_coding","Gene",12754371
chr4,129985839,129992707,6869,"+","ENSMUSG00000028773","protein_coding","Gene",59946550
chr5,136514366,136521173,6808,"+","ENSMUSG00000042985","protein_coding","Gene",21141754
chr2,113873018,113879284,6267,"-","ENSMUSG00000068614","protein_coding","Gene",15365528
chr2,172978549,172984774,6226,"+","ENSMUSG00000027513","protein_coding","Gene",74471059
chr3,10204347,10208576,4230,"-","ENSMUSG00000062515","protein_coding","Gene",46170779
chr8,112099030,112103072,4043,"-","ENSMUSG00000031722","protein_coding","Gene",87892752
chr9,53698058,53701651,3594,"+","ENSMUSG00000042045","protein_coding","Gene",18584863
chr9,110322254,110325514,3261,"+","ENSMUSG00000032484","protein_coding","Gene",75209059
chr8,126415669,126418651,2983,"-","ENSMUSG00000031972","protein_coding","Gene",102209391
chr16,48842665,48844574,1910,"+","ENSMUSG00000061100","protein_coding","Gene",37867288
chr4,147374831,147376188,1358,"+","ENSMUSG00000041616","protein_coding","Gene",77335542

XX female vs XY male: DE expressed genes with zero distance to major satellite

seqnames,"start","end","width","strand","name","biotype","Feature","DistToSat"
chr2,98502394,98504240,1847,"+","ENSMUSG00000075015","protein_coding","Gene",0
chr9,3023547,3025218,1672,"+","ENSMUSG00000074562","protein_coding","Gene",0
chr9,3017408,3019022,1615,"+","ENSMUSG00000074563","protein_coding","Gene",0
chr9,3015654,3017210,1557,"+","ENSMUSG00000074564","protein_coding","Gene",0
chr9,3020155,3021593,1439,"+","ENSMUSG00000079719","protein_coding","Gene",0
chr9,3000922,3002330,1409,"+","ENSMUSG00000091028","protein_coding","Gene",0
chr9,3025417,3026387,971,"+","ENSMUSG00000074561","protein_coding","Gene",0

chr2,98506704,98507458,755,"-","ENSMUSG00000075014","protein_coding","Gene",0

XX female vs XY male: DE genes more than1MB distant from repeats

seqnames,"start","end","width","strand","name","biotype","Feature","DistToSat"
chrX,165233031,165742367,509337,"+","ENSMUSG00000031355","protein_coding","Gene",92315149
chr2,179177136,179634078,456943,"+","ENSMUSG00000000305","protein_coding","Gene",80669646
chr2,12028259,12223547,195289,"-","ENSMUSG00000026768","protein_coding","Gene",86278843

chr5,125497523,125659589,162067,"-","ENSMUSG00000029478","protein_coding","Gene",10124911
chr6,130128771,130285823,157053,"-","ENSMUSG00000033024","protein_coding","Gene",69491446
chr4,138908524,139045503,136980,"+","ENSMUSG00000066036","protein_coding","Gene",68869235
chr6,115311239,115440419,129181,"+","ENSMUSG00000000440","protein_coding","Gene",54673914
chr3,80488858,80606713,117856,"-","ENSMUSG00000033981","protein_coding","Gene",19178303
chr5,51845488,51958964,113477,"-","ENSMUSG00000029167","protein_coding","Gene",52626541
chr2,66511482,66622971,111490,"-","ENSMUSG00000034810","protein_coding","Gene",31879419
chr12,103521258,103629065,107808,"+","ENSMUSG00000044456","protein_coding","Gene",27615765
chr11,117883037,117990533,107497,"-","ENSMUSG00000033987","protein_coding","Gene",114878284
chr5,17287508,17394777,107270,"-","ENSMUSG00000002944","protein_coding","Gene",87190728
chr4,153544839,153649822,104984,"+","ENSMUSG00000057751","protein_coding","Gene",83505550
chr11,77576748,77679482,102735,"+","ENSMUSG00000000631","protein_coding","Gene",74571995
chr4,137024634,137126545,101912,"+","ENSMUSG00000028763","protein_coding","Gene",66985345
chr9,123276058,123371782,95725,"+","ENSMUSG00000035202","protein_coding","Gene",88162863
chr2,60130160,60221360,91201,"-","ENSMUSG00000026980","protein_coding","Gene",38281030
chr2,67284059,67364671,80613,"+","ENSMUSG00000027022","protein_coding","Gene",31137719
chrX,82947276,83022158,74883,"-","ENSMUSG00000025059","protein_coding","Gene",10029394
chr13,37437214,37510905,73692,"+","ENSMUSG00000021423","protein_coding","Gene",31033224
chr9,75079821,75153258,73438,"+","ENSMUSG00000033590","protein_coding","Gene",39966626
chrX,165688125,165758304,70180,"-","ENSMUSG00000031352","protein_coding","Gene",92770243
chr5,143512088,143579341,67254,"-","ENSMUSG00000039477","protein_coding","Gene",28139476
chr8,47556395,47621405,65011,"+","ENSMUSG00000018796","protein_coding","Gene",23350117
chr2,26313423,26372183,58761,"-","ENSMUSG00000026923","protein_coding","Gene",72130207
chr6,131249162,131303615,54454,"-","ENSMUSG00000032899","protein_coding","Gene",70611837
chr3,113258628,113309617,50990,"-","ENSMUSG00000074264","protein_coding","Gene",13473310
chr13,38243197,38290446,47250,"+","ENSMUSG00000054889","protein_coding","Gene",31839207
chr11,104411977,104457067,45091,"+","ENSMUSG00000061086","protein_coding","Gene",101407224
chrX,148667563,148709078,41516,"+","ENSMUSG00000025332","protein_coding","Gene",75749681
chr3,69623460,69663818,40359,"-","ENSMUSG00000043461","protein_coding","Gene",13243786
chr3,130936337,130975019,38683,"-","ENSMUSG00000027984","protein_coding","Gene",31151019
chr5,30890543,30927643,37101,"+","ENSMUSG00000049265","protein_coding","Gene",73657862
chrX,163784254,163820681,36428,"-","ENSMUSG00000025742","protein_coding","Gene",90866372
chr2,126378143,126413979,35837,"+","ENSMUSG00000027359","protein_coding","Gene",27870653
chr16,36750264,36785048,34785,"-","ENSMUSG00000022899","protein_coding","Gene",25774887
chr6,124380777,124414812,34036,"-","ENSMUSG00000008153","protein_coding","Gene",63743452
chr2,119444034,119476980,32947,"+","ENSMUSG00000027306","protein_coding","Gene",20936544
chr11,40514890,40546939,32050,"-","ENSMUSG00000020330","protein_coding","Gene",37510137
chr11,55017852,55049832,31981,"+","ENSMUSG00000020261","protein_coding","Gene",52013099
chr8,95689967,95721737,31771,"-","ENSMUSG00000056973","protein_coding","Gene",71483689
chr6,73198376,73226905,28530,"+","ENSMUSG00000052738","protein_coding","Gene",12561051
chr5,38667211,38695448,28238,"+","ENSMUSG00000051596","protein_coding","Gene",65890057
chr18,74938851,74965861,27011,"+","ENSMUSG00000036880","protein_coding","Gene",18839550
chr8,71404390,71431347,26958,"+","ENSMUSG00000015568","protein_coding","Gene",47198112
chr11,54971972,54998579,26608,"-","ENSMUSG00000020264","protein_coding","Gene",51967219
chr3,20115516,20142080,26565,"-","ENSMUSG00000001865","protein_coding","Gene",36237275
chr6,5118090,5143946,25857,"-","ENSMUSG00000002588","protein_coding","Gene",38149813

chrX,163742861,163768490,25630,"-","ENSMUSG00000044583","protein_coding","Gene",90824979
chr6,125270677,125294961,24285,"+","ENSMUSG00000030340","protein_coding","Gene",64633352
chr18,67503218,67527448,24231,"+","ENSMUSG00000024526","protein_coding","Gene",11403917
chr3,153585321,153607596,22276,"-","ENSMUSG00000062908","protein_coding","Gene",53800003
chr9,89948990,89970927,21938,"+","ENSMUSG00000032359","protein_coding","Gene",54835795
chr6,84521938,84543740,21803,"-","ENSMUSG00000063415","protein_coding","Gene",23884613
chrX,163680626,163702261,21636,"-","ENSMUSG00000040522","protein_coding","Gene",90762744
chr2,25284223,25304060,19838,"+","ENSMUSG00000026944","protein_coding","Gene",73198330
chrX,165092049,165111830,19782,"-","ENSMUSG00000031358","protein_coding","Gene",92174167
chr2,4840065,4859776,19712,"+","ENSMUSG00000026664","protein_coding","Gene",93642614
chr16,31296278,31314894,18617,"-","ENSMUSG00000022548","protein_coding","Gene",20320901
chr4,132091465,132109657,18193,"+","ENSMUSG00000054405","protein_coding","Gene",62052176
chr9,54434288,54452468,18181,"+","ENSMUSG00000032279","protein_coding","Gene",19321093
chr6,123179861,123197039,17179,"+","ENSMUSG00000023349","protein_coding","Gene",62542536
chr11,69212010,69227177,15168,"-","ENSMUSG00000018476","protein_coding","Gene",66207257
chr6,130168623,130183340,14718,"-","ENSMUSG00000067599","protein_coding","Gene",69531298
chr19,34315580,34329826,14247,"-","ENSMUSG00000035783","protein_coding","Gene",10883496
chr12,91962994,91976878,13885,"-","ENSMUSG00000007682","protein_coding","Gene",16057501
chr2,155062269,155076013,13745,"+","ENSMUSG00000047459","protein_coding","Gene",56554779
chr6,129560341,129573882,13542,"-","ENSMUSG00000030149","protein_coding","Gene",68923016
chr19,44468941,44482199,13259,"-","ENSMUSG00000037071","protein_coding","Gene",21036857
chr6,129597514,129610707,13194,"-","ENSMUSG00000052736","protein_coding","Gene",68960189
chr6,5433351,5446309,12959,"-","ENSMUSG00000019577","protein_coding","Gene",37847450
chr5,88746337,88757842,11506,"+","ENSMUSG00000064156","protein_coding","Gene",15827663
chr11,106849943,106860855,10913,"-","ENSMUSG00000018362","protein_coding","Gene",103845190
chr6,128728503,128739233,10731,"-","ENSMUSG00000030325","protein_coding","Gene",68091178
chr2,69699619,69709903,10285,"+","ENSMUSG00000068882","protein_coding","Gene",28792487
chr11,114913231,114923491,10261,"-","ENSMUSG00000048498","protein_coding","Gene",111908478
chr2,32134336,32143595,9260,"-","ENSMUSG00000044627","protein_coding","Gene",66358795
chr8,3917657,3926844,9188,"-","ENSMUSG00000065987","protein_coding","Gene",20279102
chr8,85814247,85822355,8109,"+","ENSMUSG00000031710","protein_coding","Gene",61607969
chr5,88949537,88956916,7380,"-","ENSMUSG00000067149","protein_coding","Gene",15628589
chr19,47157961,47165115,7155,"-","ENSMUSG00000071528","protein_coding","Gene",23725877
chr5,137728747,137735694,6948,"+","ENSMUSG00000023328","protein_coding","Gene",22356135
chr4,129985839,129992707,6869,"+","ENSMUSG00000028773","protein_coding","Gene",59946550
chr9,45737711,45744141,6431,"-","ENSMUSG00000032085","protein_coding","Gene",10624516
chr4,138331650,138337961,6312,"+","ENSMUSG00000041202","protein_coding","Gene",68292361
chr2,113873018,113879284,6267,"-","ENSMUSG00000068614","protein_coding","Gene",15365528
chr4,136436044,136442102,6059,"-","ENSMUSG00000036905","protein_coding","Gene",66396755
chr6,129130633,129136552,5920,"+","ENSMUSG00000030157","protein_coding","Gene",68493308
chr2,121973422,121978819,5398,"+","ENSMUSG00000060802","protein_coding","Gene",23465932
chr11,83401384,83406589,5206,"-","ENSMUSG00000018927","protein_coding","Gene",80396631
chr13,67001553,67006314,4762,"-","ENSMUSG00000021520","protein_coding","Gene",10571855
chr13,101421298,101426014,4717,"+","ENSMUSG00000052293","protein_coding","Gene",23842979
chr5,38126758,38131059,4302,"+","ENSMUSG00000062329","protein_coding","Gene",66454446
chr11,115277003,115281276,4274,"-","ENSMUSG00000034566","protein_coding","Gene",112272250

chr3,10204347,10208576,4230,"-","ENSMUSG00000062515","protein_coding","Gene",46170779
chr6,124779565,124783719,4155,"+","ENSMUSG00000023505","protein_coding","Gene",64142240
chr2,38857100,38861144,4045,"-","ENSMUSG00000062997","protein_coding","Gene",59641246
chr8,112099030,112103072,4043,"-","ENSMUSG00000031722","protein_coding","Gene",87892752
chr12,114523536,114527141,3606,"-","ENSMUSG00000076612","protein_coding","Gene",38618043
chr4,132086470,132089574,3105,"-","ENSMUSG00000054428","protein_coding","Gene",62047181
chr4,136451832,136454718,2887,"-","ENSMUSG00000036887","protein_coding","Gene",66412543
chr4,86500564,86503316,2753,"-","ENSMUSG00000028495","protein_coding","Gene",16461275
chr8,73418621,73421342,2722,"-","ENSMUSG00000045128","protein_coding","Gene",49212343
chr11,83517493,83519770,2278,"+","ENSMUSG00000069792","protein_coding","Gene",80512740
chrX,163645025,163647247,2223,"-","ENSMUSG00000049775","protein_coding","Gene",90727143
chr11,81928687,81930301,1615,"+","ENSMUSG00000009185","protein_coding","Gene",78923934
chr13,23855537,23856283,747,"+","ENSMUSG00000049539","protein_coding","Gene",17451547
chr13,23627291,23627820,530,"+","ENSMUSG00000062417","protein_coding","Gene",17223301

XX intact female vs XX gonadectomised female: DE expressed genes with zero distance to major satellite

chr2,98502394,98504240,1847,"+","ENSMUSG00000075015","protein_coding","Gene",0
chr9,3023547,3025218,1672,"+","ENSMUSG00000074562","protein_coding","Gene",0
chr9,3017408,3019022,1615,"+","ENSMUSG00000074563","protein_coding","Gene",0
chr9,3020155,3021593,1439,"+","ENSMUSG00000079719","protein_coding","Gene",0
chr9,3000922,3002330,1409,"+","ENSMUSG00000091028","protein_coding","Gene",0
chr2,98506704,98507458,755,"-","ENSMUSG00000075014","protein_coding","Gene",0

XX intact female vs XX gonadectomised female: DE genes more than1MB distant from repeats

chr3,80488858,80606713,117856,"-","ENSMUSG00000033981","protein_coding","Gene",19178303
chr19,4510472,4621752,111281,"+","ENSMUSG00000024892","protein_coding","Gene",18810159
chr4,137024634,137126545,101912,"+","ENSMUSG00000028763","protein_coding","Gene",66985345
chr9,123276058,123371782,95725,"+","ENSMUSG00000035202","protein_coding","Gene",88162863
chr2,67284059,67364671,80613,"+","ENSMUSG00000027022","protein_coding","Gene",31137719
chr4,102710670,102787160,76491,"-","ENSMUSG00000035126","protein_coding","Gene",32671381
chr8,47556395,47621405,65011,"+","ENSMUSG00000018796","protein_coding","Gene",23350117
chrX,160414754,160465929,51176,"-","ENSMUSG00000031373","protein_coding","Gene",87496872
chr2,152673700,152721057,47358,"+","ENSMUSG00000027469","protein_coding","Gene",54166210
chr16,36750264,36785048,34785,"-","ENSMUSG00000022899","protein_coding","Gene",25774887
chr8,71404390,71431347,26958,"+","ENSMUSG00000015568","protein_coding","Gene",47198112
chr11,66984529,67011019,26491,"+","ENSMUSG00000033196","protein_coding","Gene",63979776
chr11,105869438,105895863,26426,"+","ENSMUSG00000001901","protein_coding","Gene",102864685
chr11,67051314,67073948,22635,"+","ENSMUSG00000057003","protein_coding","Gene",64046561
chr9,72286336,72306381,20046,"+","ENSMUSG00000032221","protein_coding","Gene",37173141

chr2,164665018,164683211,18194,"-","ENSMUSG00000017754","protein_coding","Gene",66157528
chr6,138088837,138105276,16440,"+","ENSMUSG00000008540","protein_coding","Gene",77451512
chr4,149561124,149575441,14318,"-","ENSMUSG00000028972","protein_coding","Gene",79521835
chrX,130683582,130696615,13034,"-","ENSMUSG00000031258","protein_coding","Gene",57765700
chr5,88746337,88757842,11506,"+","ENSMUSG00000064156","protein_coding","Gene",15827663
chr16,23146609,23158101,11493,"+","ENSMUSG00000022878","protein_coding","Gene",12171232
chr11,101191919,101203252,11334,"+","ENSMUSG00000019326","protein_coding","Gene",98187166
chr6,113374628,113385754,11127,"-","ENSMUSG00000030278","protein_coding","Gene",52737303
chr3,14863538,14872351,8814,"+","ENSMUSG00000027559","protein_coding","Gene",41507004
chr11,54715955,54723879,7925,"+","ENSMUSG00000018339","protein_coding","Gene",51711202
chr2,113873018,113879284,6267,"-","ENSMUSG00000068614","protein_coding","Gene",15365528
chr2,172978549,172984774,6226,"+","ENSMUSG00000027513","protein_coding","Gene",74471059
chr2,121973422,121978819,5398,"+","ENSMUSG00000060802","protein_coding","Gene",23465932
chr8,3655770,3660110,4341,"+","ENSMUSG00000012705","protein_coding","Gene",20545836
chr3,10204347,10208576,4230,"-","ENSMUSG00000062515","protein_coding","Gene",46170779
chr8,87604639,87606238,1600,"+","ENSMUSG00000059355","protein_coding","Gene",63398361
chr12,104680377,104681890,1514,"-","ENSMUSG00000079017","protein_coding","Gene",28774884
chr6,72907341,72908742,1402,"-","ENSMUSG00000079523","protein_coding","Gene",12270016
chr11,32183511,32184465,955,"+","ENSMUSG00000069919","protein_coding","Gene",29178758
chr11,32196489,32197301,813,"+","ENSMUSG00000069917","protein_coding","Gene",29191736

XX intact female vs XY gonadectomised male: DE expressed genes with zero distance to major satellite

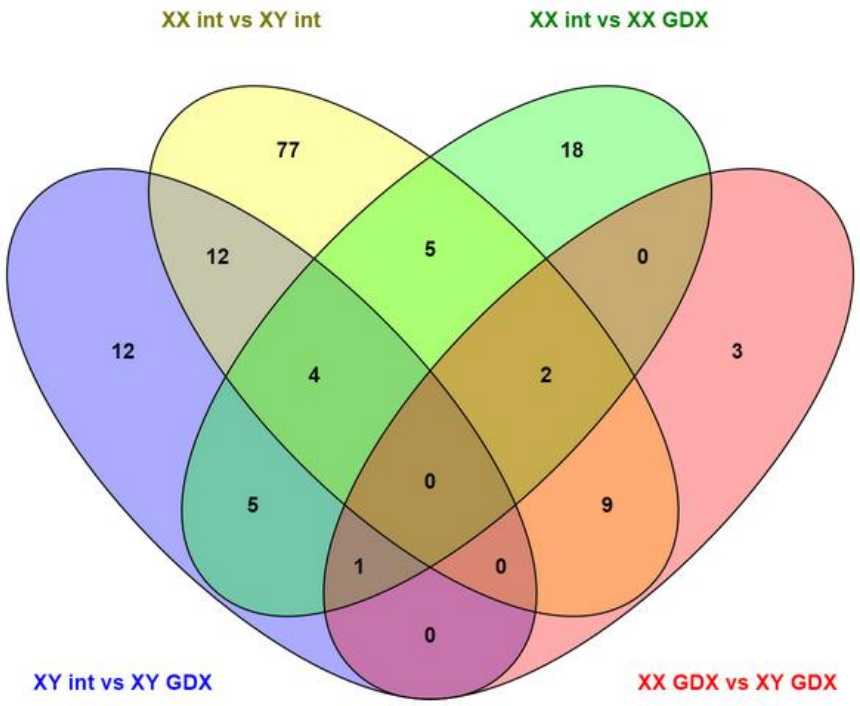
chr2,98502394,98504240,1847,"+","ENSMUSG00000075015","protein_coding","Gene",0
chr9,3023547,3025218,1672,"+","ENSMUSG00000074562","protein_coding","Gene",0
chr9,3017408,3019022,1615,"+","ENSMUSG00000074563","protein_coding","Gene",0
chr9,3015654,3017210,1557,"+","ENSMUSG00000074564","protein_coding","Gene",0
chr9,3020155,3021593,1439,"+","ENSMUSG00000079719","protein_coding","Gene",0
chr9,3000922,3002330,1409,"+","ENSMUSG00000091028","protein_coding","Gene",0
chr9,3034599,3035805,1207,"+","ENSMUSG00000091057","protein_coding","Gene",0
chr9,3037111,3038316,1206,"+","ENSMUSG00000074558","protein_coding","Gene",0
chr9,3013170,3014344,1175,"+","ENSMUSG00000074565","protein_coding","Gene",0
chr9,3025417,3026387,971,"+","ENSMUSG00000074561","protein_coding","Gene",0
chr2,98506704,98507458,755,"-","ENSMUSG00000075014","protein_coding","Gene",0
chr9,3032085,3032822,738,"+","ENSMUSG00000061971","protein_coding","Gene",0
chr9,3030337,3031071,735,"+","ENSMUSG00000091599","protein_coding","Gene",0

XX intact female vs XX gonadectomised female: DE genes more than1MB distant from repeats

chr4,102413902,102646233,232332,"+","ENSMUSG00000028524","protein_coding","Gene",32374613
chr6,130128771,130285823,157053,"-","ENSMUSG00000033024","protein_coding","Gene",69491446
chrX,17739701,17857062,117362,"+","ENSMUSG00000037369","protein_coding","Gene",55032622
chr9,123276058,123371782,95725,"+","ENSMUSG00000035202","protein_coding","Gene",88162863
chrX,165688125,165758304,70180,"-","ENSMUSG00000031352","protein_coding","Gene",92770243
chrX,163784254,163820681,36428,"-","ENSMUSG00000025742","protein_coding","Gene",90866372

chr16,36750264,36785048,34785,"-","ENSMUSG00000022899","protein_coding","Gene",25774887
chrX,163742861,163768490,25630,"-","ENSMUSG00000044583","protein_coding","Gene",90824979
chrX,163680626,163702261,21636,"-","ENSMUSG00000040522","protein_coding","Gene",90762744
chrX,165092049,165111830,19782,"-","ENSMUSG00000031358","protein_coding","Gene",92174167
chr6,29010220,29023877,13658,"+","ENSMUSG00000059201","protein_coding","Gene",14269882
chrX,130683582,130696615,13034,"-","ENSMUSG00000031258","protein_coding","Gene",57765700
chr5,88949537,88956916,7380,"-","ENSMUSG00000067149","protein_coding","Gene",15628589
chr12,114523536,114527141,3606,"-","ENSMUSG00000076612","protein_coding","Gene",38618043
chrX,163645025,163647247,2223,"-","ENSMUSG00000049775","protein_coding","Gene",90727143

6.1.1 DE genes not proximal to repeats



6.2 Cell concentrations in human and mouse samples

Major cell populations	Human PBMC* (typical number/L)	Mouse spleen (typical number/spleen)	Mouse lymph nodes
Red blood cells	NA†	~2.5–5 x 10^{10}	Variable
Platelets	NA†	NA†	NA†
White blood cells	~1–2 x 10^{9} per L human blood	~1 x 10^{8}	Variable
Typical % distribution in white blood cells			
T cells	40–60%	30–35%	60–70%
$CD4^+$ T cells	~70% of T cells	~70% of T cells	~70% of T cells
$CD8^+$ T cells	~30% of T cells	~30% of T cells	~30% of T cells
B cells	3–15%	45–50%	30%
NK cells	~10%	~10%	Rare
Neutrophils	NA†	5–10%	<5%
Monocytes	15–35%		
Eosinophils	NA†		
Basophils	NA†		

* PBMC, peripheral blood mononuclear cells. † NA = Not applicable.