



Published in final edited form as:

*Nat Genet.* 2008 June ; 40(6): 794–799. doi:10.1038/ng.126.

## The mouse X chromosome is enriched for multi-copy testis genes exhibiting post-meiotic expression

Jacob L. Mueller<sup>1</sup>, Shantha K. Mahadevaiah<sup>2</sup>, Peter J. Park<sup>3</sup>, Peter E. Warburton<sup>4</sup>, David C. Page<sup>1</sup>, and James M.A. Turner<sup>2</sup>

<sup>1</sup> Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

<sup>2</sup> Division of Stem Cell Research and Developmental Genetics, MRC National Institute for Medical Research, London NW7 1AA, United Kingdom

<sup>3</sup> Children's Hospital Informatics Program, Harvard-Partners Center for Genetics and Genomics, Boston, Massachusetts 02115

<sup>4</sup> Department of Human Genetics, Mount Sinai School of Medicine, New York, New York 10029, USA

### Abstract

According to the prevailing view, mammalian X chromosomes are enriched in spermatogenesis genes expressed before meiosis<sup>1–3</sup> and deficient in spermatogenesis genes expressed after meiosis<sup>2,3</sup>. The paucity of post-meiotic genes on the X chromosome has been interpreted as a consequence of Meiotic Sex Chromosome Inactivation (MSCI) – the complete silencing of genes on the XY bivalent at meiotic prophase<sup>4,5</sup>. Recent studies have concluded that MSCI-initiated silencing persists beyond meiosis<sup>6–8</sup> and that most X-genes remain repressed in round spermatids<sup>7</sup>. We report here that 33 multi-copy gene families, representing ~273 mouse X-linked genes, are expressed in the testis and that this expression is predominantly in post-meiotic cells. RNA FISH and microarray analysis show that the maintenance of X chromosome post-meiotic repression is incomplete. Furthermore, X-linked multi-copy genes exhibit expression levels similar to those of autosomal genes. Thus, not only is the mouse X chromosome enriched for spermatogenesis genes functioning before meiosis, but in addition ~18% of mouse X-linked genes exhibit post-meiotic expression.

---

The existing model that the X chromosome is deficient in post-meiotic spermatogenesis genes was based primarily on the analysis of single-copy genes<sup>2,3</sup>. We therefore sought to explore the spermatogenesis expression profiles of X-linked multi-copy genes. We performed a systematic search for ampliconic regions - comprising palindromic or tandem

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

#### Author contributions

Amplicon identification and multicopy gene discovery were performed by JLM, JMAT, SKM and PEW. RT-PCRs and RNA FISH were performed by JLM, JMAT and SKM. Microarray and statistical analysis were performed by JLM and PJP. The paper was written by JLM, JMAT and DCP.

segmental duplications - and their associated genes on the mouse X chromosome. To identify palindromic repeats, we used the Inverted Repeats Finder (IRF) program<sup>9</sup>, focusing on palindromes with large arms (> 8kb), exhibiting > 90% nucleotide identity between arms, and are <500kb from each other. Using these criteria, we identified 17 palindromic regions, seven of which contained multiple types of repeat units (Table 1). In order to detect amplicons consisting only of tandem repeats, we searched for multi-copy gene clusters (Supplementary Table 1). This search revealed five additional ampliconic regions (Amp7, Amp17, Amp19, Amp21, and Amp22; Table 1), and confirmed the existence of all 17 IRF-identified palindromes. In sum, we identified 22 ampliconic regions consisting of 29 distinct repeat units whose sequence complexity totals 1,474kb (Table 1). Together, these ampliconic sequences comprise approximately 19.4Mb, or ~12% of the ~166Mb mouse X chromosome (Table 1).

The contiguous assembly of ampliconic regions is known to be difficult<sup>10</sup>. Of the 24 physical gaps in the mouse X chromosome (NCBI *Mus musculus* genome Build 37.1), 21 reside within amplicons (Supplementary Table 2). Five regions of the mouse X chromosome have yet to be completely assembled, because they contain huge (>2Mb) and highly complex ampliconic structures (Amp1, Amp4, Amp7, Amp17, and Amp19; Fig. 1a, b; Table 1). These five regions represent the vast majority (an estimated 87% or 16.9Mb, Supplementary Table 2) of the 19.4Mb of ampliconic sequence and most (19 of 24) X chromosome assembly gaps (Supplementary Table 2).

We next asked whether mouse X-ampliconic genes exhibit testis-biased expression. We searched X-ampliconic regions for ESTs and novel/predicted genes and identified 26 multi-copy genes located entirely within these ampliconic regions (Supplementary Table 3). RT-PCR revealed that 23 of the 26 X-ampliconic multi-copy genes are expressed predominantly or exclusively in the testis (Fig. 2). The 23 testis-expressed multi-copy genes reside in 20 of the 22 ampliconic regions (Fig. 1b, c). Examination of the genomic structure of these 20 regions (Supplementary Fig. 1) showed that the X-ampliconic gene copy number ranges from two to ~28 (Fig. 1c), with a total of ~232 ampliconic protein-coding gene copies exhibiting testis-biased expression.

To determine if testis-biased expression of multi-copy genes is a phenomenon limited to ampliconic regions, we also analyzed non-ampliconic multi-copy genes. We defined all regions of the mouse X chromosome not harboring ampliconic sequences, as non-ampliconic sequence (~88% of the mouse X chromosome). We selected genes present in at least two copies, located within 1Mb of each other and with > 85% amino acid identity (Supplementary Table 1). Ten multi-copy genes were identified (Fig. 1c), which ranged in copy number from two to 14 and totaling to 41 gene copies. Like the X-ampliconic genes, all ten multi-copy genes exhibited testis-biased expression (Fig. 2). This indicates that the mouse X chromosome has at least 33 multi-copy genes (23 ampliconic and 10 non-ampliconic) with testis-biased gene expression. Considering there are 1,555 X-linked genes (NCBI Build 37.1), then ~18% (232 ampliconic + 41 non-ampliconic = 273 gene copies) of X-linked genes are multi-copy genes with testis-biased expression, assuming all copies of each family are expressed. Testis ESTs for multi-copy family members (Supplementary

Table 3) and copy-specific RT-PCR assays (Supplementary Fig. 2) both suggest that most, if not all, of the 33 multi-copy genes are expressed from more than a single copy.

We then determined whether the 33 multi-copy X-linked genes are expressed in germ cells or somatic testis cells. We performed RT-PCR on *XXSry* and *XXSxr<sup>b</sup>* testes that lack germ cells due to a combination of the absence of spermatogonial proliferation factor *Eif2s3y* as well as the toxic effects of double X chromosome dosage<sup>11,12</sup>. For 28 of the 33 multi-copy genes, no expression was detected in either *XXSry* or *XXSxr<sup>b</sup>* testes (Fig. 3a), indicating that these 28 multi-copy genes are germ cell-specific.

Next, we established at what stage of spermatogenesis these 28 germ cell-specific genes initiate expression. In prepubertal mice, the first wave of spermatogenesis is a synchronized process, with progressively more mature spermatogenic cell types appearing at defined time points after birth. Prepubertal mice thus allow the correlation of expression patterns with the appearance of specific germ cell substages. We found that 20 of the 28 multi-copy X-linked genes initiate expression at or after 18.5dpp (Fig. 3b), when secondary spermatocytes and round spermatids first appear. The other eight X-linked genes initiate expression before meiosis, at 7.5dpp (Fig. 3b), and this early expression would mask any potential post-meiotic reactivation (Fig. 4). Despite this, we find that the majority (20 of 28) of X-linked multi-copy genes are specifically expressed in post-meiotic spermatogenic cells. This finding is unexpected since the X chromosome retains a repressed transcriptional state in post-meiotic cells<sup>6–8</sup>.

To address the degree of post-meiotic repression, we examined nascent transcription of individual X-linked genes in round spermatids via RNA FISH (Fig. 4). We selected four single-copy X-linked genes: *Birc4*, *Scml2*, *Fmr1* and *Zfx*. In a previous microarray study, each of these genes was found to be subject to MSCI and to remain repressed during spermiogenesis<sup>7</sup>. To discriminate between X- and Y-bearing round spermatids we included a second RNA FISH probe against the spermatid-expressed Y-linked gene *Sly* (Supplementary Fig. 3a), which marks all Y-bearing round spermatids (data not shown). To ensure that the efficiency of RNA FISH was equivalent between different spermatogenic cell types, we also performed RNA FISH on mice carrying a ubiquitously expressed autosomal transgene (Supplementary Fig. 3b). All four X-linked genes are expressed in spermatogonia and are then silenced in all late pachytene spermatocytes examined (Fig. 4; see legend for quantitation). This repression is maintained in the majority of X-bearing round spermatids but in contrast to previous reports, three of the four genes also exhibit some degree of reactivation: 0%, 7%, 17% and 18% for *Birc4*, *Scml2*, *Fmr1* and *Zfx*, respectively (see legend of Fig. 4 and Supplementary Table 4 for details of quantification).

Do other single-copy X-linked genes behave the same way? We analyzed published microarray data from isolated germ cell populations to examine the post-meiotic expression levels of 278 single-copy genes classified as being expressed specifically in A or B spermatogonia<sup>7</sup>. We calculated the average expression values for these 278 genes in pachytene spermatocytes and round spermatids and found that round spermatids exhibit a significantly higher average level of expression than pachytene spermatocytes ( $P < 0.0002$ , Wilcoxon Ranks Sum, Fig. 5a). This suggests there is a general low level of post-meiotic

reactivation for many genes that are also expressed earlier in A or B spermatogonia. We conclude that, in contrast to the complete meiotic silencing of X-linked genes during MSCI, post-meiotic repression of the X chromosome is incomplete.

The reactivation of single-copy X-linked genes in a small percentage of round spermatids provides a potential explanation for the amplification of spermatid expressed X-linked genes; increasing gene copy number raises the probability that an X-bearing round spermatid will express a given multi-copy gene. To test this we examined three multi-copy genes by RNA FISH; *Ott* (~12 copies), *Slx like-1* (~14 copies) and *Slx* (~25 copies). Unlike the single-copy X-linked genes, the majority of X-bearing round spermatids showed RNA-FISH signals for all three multi-copy genes (Fig. 4). Furthermore, the percentage of X-bearing round spermatids that exhibited expression increased with gene copy number; 54% for *Ott*, 78% for *Slx like-1* and 93% for *Slx*. For comparison, we examined the frequency of round spermatid expression for a single-copy spermatid-specific autosomal gene, *Zfp29*. We detected *Zfp29* RNA FISH signals in 64% of round spermatids; a frequency similar to the three X-linked multi-copy genes. The high frequency, as compared to single-copy X-linked genes, of X-bearing round spermatids expressing a given multi-copy gene suggests that gene amplification facilitates higher levels of gene expression in the face of post-meiotic X chromosome repression.

Do multi-copy X-linked genes show higher spermatid expression levels than single-copy X-linked genes at the level of mature RNA? To address this we analyzed published microarray data from isolated cell populations of pachytene spermatocytes and round spermatids<sup>7</sup>. We identified 24 X-linked multi-copy genes, 552 X-linked single-copy genes and 20,236 autosomal genes. In pachytene spermatocytes, both X-linked multi-copy and single-copy genes exhibit expression levels consistent with the presence of MSCI (Fig. 5b). However, in round spermatids X-linked multi-copy genes show a significantly higher average level of expression than X-linked single-copy genes ( $P < 10^{-5}$ , Wilcoxon Ranks Sum, Fig. 5b; Supplementary Table 5). This high level of post-meiotic X-linked multi-copy gene expression further suggests that X-gene amplification compensates for post-meiotic repression.

Evolutionary models predict the accumulation of male advantage alleles on X chromosomes<sup>13</sup>. X-linked spermatogenesis genes identified prior to this study are primarily expressed in pre-meiotic cells<sup>1,2</sup>. Here we describe a collection of 33 X-linked multi-copy genes that are expressed in post-meiotic spermatogenic cells. Some of these gene families may have important roles during sperm maturation, including *Mgclh*, whose autosomal paralogue *Mgcl1* is required for acrosome formation and spermatid chromatin condensation<sup>14</sup> and *1700012L04Rik*, which encodes a novel spermatid-specific histone variant, H2AL115.

To understand the unusual expression pattern of X-linked multi-copy genes in the face of post-meiotic repression, we have re-examined the extent of X chromosome repression during spermiogenesis. Post-meiotic X-repression is incomplete, with pre-meiotic expressed single-copy X-linked genes exhibiting reactivation at varying levels. This reconciles the apparent conflict between studies documenting the existence of a repressed X chromosome

in post-meiotic spermatogenic cells<sup>6–8</sup> and those showing reactivation of selected single-copy X-linked genes<sup>16</sup>. Importantly, we find that X-linked multi-copy genes yield an elevated frequency of expressing X-bearing round spermatids and the average expression levels of these genes is higher than X-linked single-copy genes. It is possible that other compensatory mechanisms, aside from increased copy number, also counteract post-meiotic repression, because rare cases of robust reactivation of single-copy X-linked genes (*e.g.* *Ube1x*) have also been uncovered<sup>8</sup>. Nevertheless, amplification of X-linked genes may have evolved to compensate for the repressive chromatin environment affecting the X chromosome in post-meiotic cells (Fig. 5c). Indeed, the human X chromosome harbors multi-copy genes similar to those described here as well as primate-specific multi-copy genes exhibiting post-meiotic expression<sup>17,9,18</sup>.

## METHODS

### Amplicon identification and multi-copy gene discovery

Individual mouse X chromosome amplicons (palindromic or tandem segmental duplications) were first identified using the Inverted Repeats Finder (IRF) program<sup>9</sup>. The *Mus musculus* reference sequence (NCBI Build 36.1) was used to select palindromes with arms  $\geq 8$ kb,  $\geq 90\%$  nucleotide identity between arms, and  $<500$ kb of each other. The  $\geq 8$ kb arm size restriction, used in the Warburton et al. (2004) palindrome analyses, should eliminate all repeats due to recent LINE insertions. Custom tracks can be downloaded from <https://tandem.bu.edu> to display all mouse X chromosome palindromes on the UCSC Genome Browser. In some cases the IRF program truncated and/or missed palindrome arms. Therefore, to resolve the complete structure of all IRF-identified palindromes, custom perl code<sup>19</sup> was used to characterize their genomic structure via triangular dot-plots (Supplementary Fig. 1).

Since the IRF program identifies only palindromes, we utilized an alternative strategy to detect amplicons in tandem, which could also confirm the IRF-identified amplicons. Ensembl's Biomart program (<http://www.ensembl.org/>) was used to obtain all X-linked protein-coding genes with paralogous genes on the X chromosome (Supplementary Table 1). Multi-copy genes less than 1Mb apart, which share  $\geq 80\%$  amino acid identity between any two copies, and which are not found in multiple copies throughout the mouse genome (*e.g.* retroviral proteins, ribosomal proteins, or olfactory receptors) were considered putative ampliconic regions. Multi-copy gene clusters passing these three criteria (Supplementary Table 1) were subjected to triangular dot-plot analysis<sup>19</sup> to determine if the region is ampliconic. Ampliconic regions identified via this method confirmed the IRF-identified amplicons as well as five new tandemly arrayed amplicons (Table 1). Repeat unit boundaries and percent identities between exclusively tandemly arrayed amplicons were determined via ClustalW<sup>20</sup>. Multi-copy gene clusters fulfilling the three selection criteria, but not present in ampliconic regions were termed non-ampliconic multi-copy genes. To estimate gene copy number, we used both the NCBI Build 37.1 annotation and protein sequence searches via TBLASTN to the mouse X chromosome reference sequence.

To identify all protein-coding genes within the ampliconic repeats, each repeat unit was repeat masked (<http://www.repeatmasker.org>) and subsequently compared to the NCBI

mouse EST database using BLAST21. We also searched, via BLAST, for predicted genes without EST support (NCBI mouse *ab initio* database) to uncover novel transcripts. Only genes falling entirely within a repeat unit were considered candidate ampliconic genes. Supplementary Table 3 provides a list of all novel and known genes identified via BLAST searches.

### RT-PCR and RNA/DNA FISH

Total RNA was extracted using Trizol® (Gibco BRL) according to manufacturer's instructions. For DNase treatment, 3µg total RNA was incubated with 1.5 U RQ DNase 1 (Promega) at 37°C for 90 mins. The RNA was precipitated, resuspended and then reverse transcribed using Superscript II reverse transcriptase (200 U, Gibco BRL) for 90 mins at 42°C. PCR was performed using the parameters: 1 cycle 94°C 3 mins, 35 cycles 94°C 30secs, 56°C 30secs, 72°C 30secs, and then 1 cycle 72°C 10mins. Primer sequences (Supplementary Table 3) for all RT-PCR reactions were designed to detect at least two copies of each family member. RNA/DNA FISH was carried out exactly as described in Turner et al. (2006). We used long range PCR products from BACs as probes for *Birc4*, *Fmr1*, *Scml2*, *Ott*, *Slx-like 1*, *Slx*, and *Zfx*. Primer sequences and BAC identifier names are in Supplementary Table 3.

### Microarray and Statistical Analyses

We analyzed published microarray expression data (Gene Expression Omnibus GSE4193)7 from isolated germ cell populations of A spermatogonia, B spermatogonia, pachytene spermatocytes and round spermatids (two replicates for each cell population). Each array was normalized as in Namekawa et al. (2006) to set the trimmed (2% from each side) mean signal intensity to 125. Probes with <100 signal intensity at all timepoints were discarded, because their expression level estimates are unreliable. Probes which mapped to autosomal genes were also removed from analysis. Single-copy X-linked genes, summing to 552, were selected from Ensembl annotations of genes without mouse paralogs. The 25 multi-copy genes' probe-ids and expression values for each timepoint are listed in Supplementary Table 5. When multiple probes matched a multi-copy gene (see Supplementary Table 5 for genes with multiple probe\_ids) we selected a single probe, with the median signal intensity. Wilcoxon Ranks Sum paired tests, using JMP 5.1 software (SAS Institute Inc.), were performed on all comparisons of average expression levels and were used because they diminish the influence of outliers.

### Supplementary Material

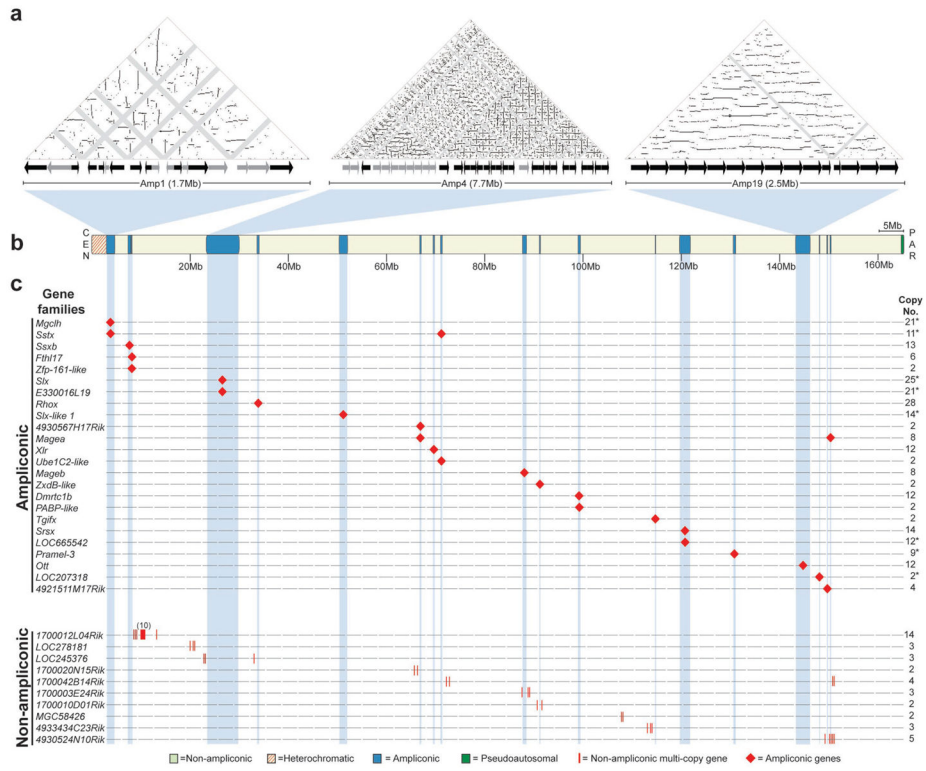
Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We thank J. Alfoldi, D. Bellott, P. Burgoyne, H. Byers, J. Hughes, J. Lange, L. Reynard, S. Rozen, H. Skaletsky for helpful comments on the manuscript; D. Bellott, M. Gill, Y. Hu, J. Hughes, H. Skaletsky for technical advice; Paul Burgoyne and Mike Mitchell for the *Uty* transgenic line. This work was supported by National Institutes of Health Fellowship F32HD052379 (to J.L.M.) and the Howard Hughes Medical Institute.

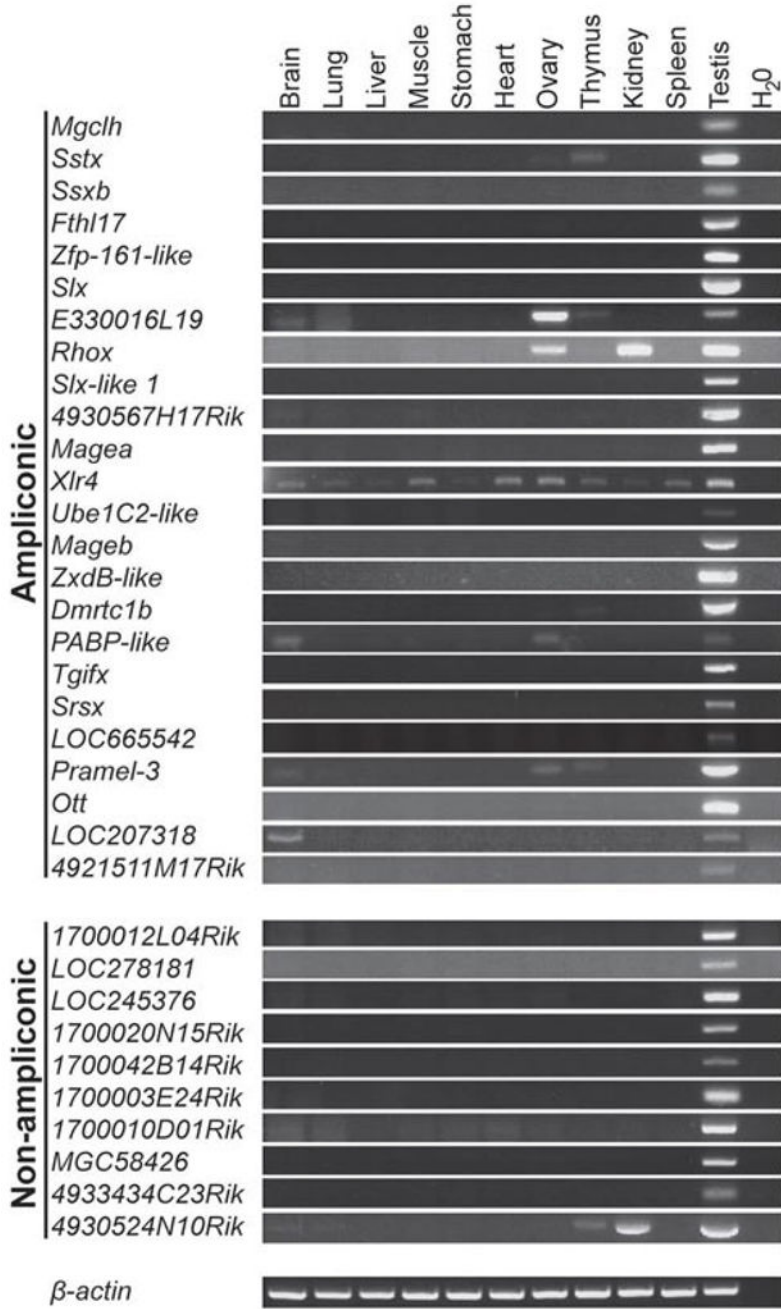
## References

1. Wang PJ, McCarrey JR, Yang F, Page DC. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet.* 2001; 27:422–6. [PubMed: 11279525]
2. Khil PP, Smirnova NA, Romanienko PJ, Camerini-Otero RD. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat Genet.* 2004; 36:642–6. [PubMed: 15156144]
3. Reinke V. Sex and the genome. *Nat Genet.* 2004; 36:548–9. [PubMed: 15167926]
4. McKee BD, Handel MA. Sex chromosomes, recombination, and chromatin conformation. *Chromosoma.* 1993; 102:71–80. [PubMed: 8432196]
5. Turner JM. Meiotic sex chromosome inactivation. *Development.* 2007; 134:1823–31. [PubMed: 17329371]
6. Greaves IK, Rangasamy D, Devoy M, Marshall Graves JA, Tremethick DJ. The X and Y chromosomes assemble into H2A. Z-containing facultative heterochromatin following meiosis. *Mol Cell Biol.* 2006; 26:5394–405. [PubMed: 16809775]
7. Namekawa SH, et al. Postmeiotic sex chromatin in the male germline of mice. *Curr Biol.* 2006; 16:660–7. [PubMed: 16581510]
8. Turner JM, Mahadevaiah SK, Ellis PJ, Mitchell MJ, Burgoyne PS. Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial postmeiotic repression in spermatids. *Dev Cell.* 2006; 10:521–9. [PubMed: 16580996]
9. Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res.* 2004; 14:1861–9. [PubMed: 15466286]
10. Eichler EE, Clark RA, She X. An assessment of the sequence gaps: unfinished business in a finished human genome. *Nat Rev Genet.* 2004; 5:345–54. [PubMed: 15143317]
11. Mroz K, Carrel L, Hunt PA. Germ cell development in the XXY mouse: evidence that X chromosome reactivation is independent of sexual differentiation. *Dev Biol.* 1999; 207:229–38. [PubMed: 10049577]
12. Mazeyrat S, et al. A Y-encoded subunit of the translation initiation factor Eif2 is essential for mouse spermatogenesis. *Nat Genet.* 2001; 29:49–53. [PubMed: 11528390]
13. Rice WR. Sex-Chromosomes and the Evolution of Sexual Dimorphism. *Evolution.* 1984; 38:735–742.
14. Kimura T, et al. Mouse germ cell-less as an essential component for nuclear integrity. *Mol Cell Biol.* 2003; 23:1304–15. [PubMed: 12556490]
15. Govin J, et al. Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J Cell Biol.* 2007; 176:283–94. [PubMed: 17261847]
16. Wang PJ, Page DC, McCarrey JR. Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse. *Hum Mol Genet.* 2005; 14:2911–8. [PubMed: 16118233]
17. Westbrook VA, et al. Spermatid-specific expression of the novel X-linked gene product SPAN-X localized to the nucleus of human spermatozoa. *Biol Reprod.* 2000; 63:469–81. [PubMed: 10906052]
18. Westbrook VA, et al. Hominoid-specific SPANXA/D genes demonstrate differential expression in individuals and protein localization to a distinct nuclear envelope domain during spermatid morphogenesis. *Mol Hum Reprod.* 2006; 12:703–16. [PubMed: 17012309]
19. Kuroda-Kawaguchi T, et al. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet.* 2001; 29:279–86. [PubMed: 11687796]
20. Waterston RH, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 2002; 420:520–62. [PubMed: 12466850]
21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215:403–10. [PubMed: 2231712]

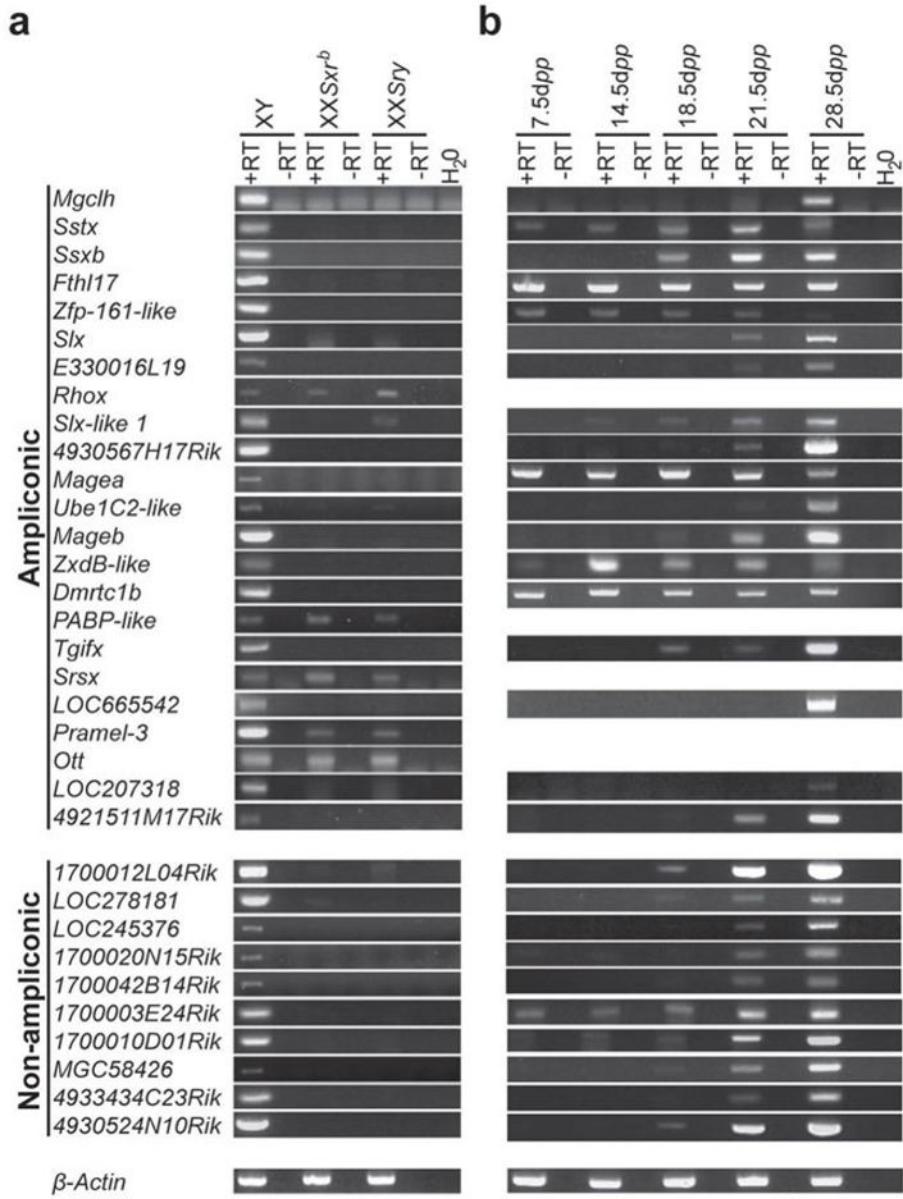


**Figure 1.** Mouse X chromosome ampliconic regions containing testis-expressed genes. **a**, Examples of complexity (Amp1, left), massive scale (Amp4, center), and tandem duplications (Amp19, right) in ampliconic regions. Each ampliconic region is compared to itself in a triangular dot-plot. Individual dots represent a perfect match of 200 nucleotides. Horizontal lines and vertical lines depict direct and inverted repeats, respectively. Shaded grey regions represent physical gaps in the sequence assembly. Primary (black) and secondary (grey) amplicons, or repeat units, are shown below the dot-plots as arrows. The size of each region is shown below the amplicon arrows. **b**, Mouse X chromosome with ampliconic regions shown in blue. **c**, Position of ampliconic and non-ampliconic multi-copy genes and their copy number. Asterisks denote copy number estimates since they fall within incompletely assembled genomic regions. Copy number estimates are based upon the number of intact ORFs in the current NCBI Build 37.1.

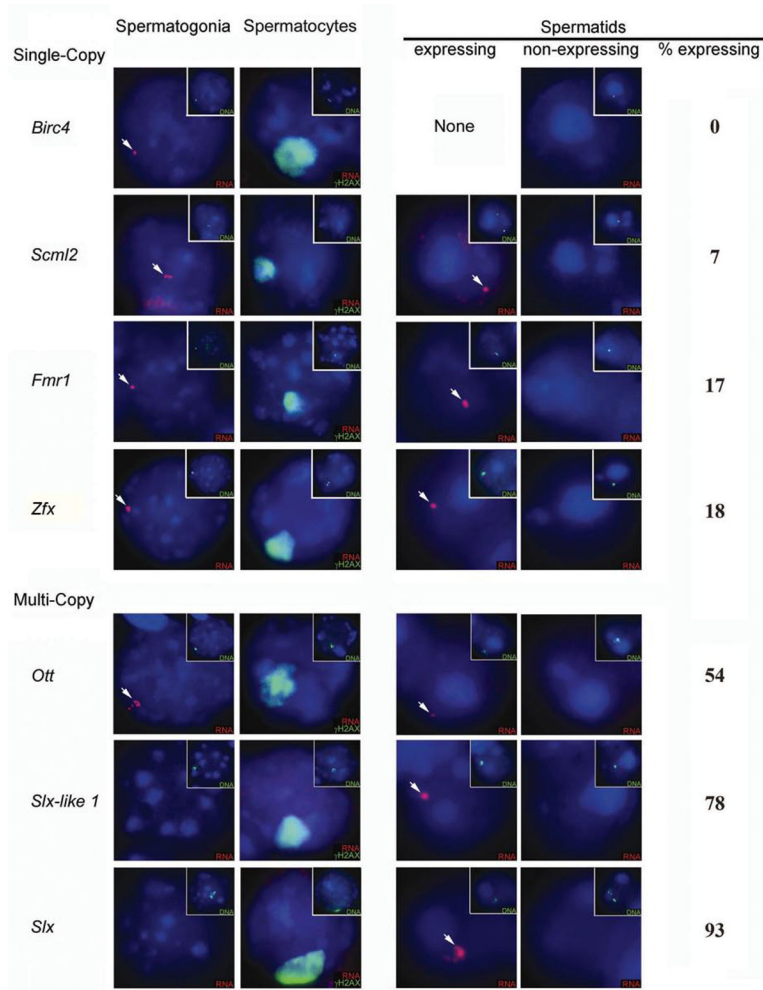




**Figure 2.** Mouse X chromosome ampliconic and non-ampliconic multi-copy genes exhibit testis-biased expression as shown by RT-PCR. We assayed 26 ampliconic and 10 non-ampliconic multi-copy genes across 11 different tissues – *Cxx* and *Obp1* were not detected in any tissues and are therefore not shown. Of the 35 multi-copy genes expressed in testis, only *Xlr4* did not exhibit testis-biased expression. Other *Xlr* gene family members (*Xlr3* and *Xlr5*) exhibit ubiquitous expression patterns similar to *Xlr4* (data not shown).  $\beta$ -Actin is ubiquitously expressed and serves as a control.



**Figure 3.** Mouse X chromosome ampliconic and non-ampliconic multi-copy genes are expressed predominantly in germ cells during post-meiotic spermatogenesis as shown by RT-PCR. **a.** All 23 multi-copy genes exhibiting testis-biased expression were assayed in wild type and germ cell negative XXSry and XXSxr<sup>b</sup> testes. Only *Rhox*, *PABP-like*, *Srsx*, *Pramel-3* and *Ott* show expression in both XXSry and XXSxr<sup>b</sup> testes. We note that *Pramel-3* gives faint products, consistent with the findings of Wang *et al.* (2001) that the predominant site of expression for this multi-copy gene is in germ cells. **b.** Genes exhibiting strict germ cell-specificity were assayed by analysis of the first wave of spermatogenesis in juvenile male mice. Twenty of the 28 genes are expressed from 18.5dpp, the remaining eight – *Sstx*, *Fthl17*, *Zfp161-like*, *Magea*, *ZxdB-like*, *Dmrtc1b*, *170003E24Rik* and *1700010D01Rik* are expressed from 7.5dpp.

**Figure 4.**

Multi-copy ampliconic genes exhibit higher levels of reactivation in round spermatids than single-copy genes. Four single-copy and three multi-copy X-linked genes were assayed for expression during successive stages of spermatogenesis by RNA/DNA FISH. First column – spermatogonia. All four single-copy genes are expressed at this stage, consistent with previously reported observations (see below for details of quantitation)<sup>7</sup>. In contrast, only one multi-copy gene, *Ott*, is expressed. Second column – pachytene spermatocytes. The sex body is shown by  $\gamma$ H2AX staining. All seven genes are silent in 100% of cells examined (n=100). Third and fourth columns represent examples of expressing and non-expressing round spermatid cells, respectively. Single-copy genes show reactivation in a minority of round spermatids, as shown by the percentages in the final column (see also Supplementary Table 4). *Birc4* expression could not be detected in any round spermatids, and so the ‘spermatid-expressing’ column is blank for this gene. In contrast, multi-copy genes show expression in a high percentage of round spermatids. Note that although *Ott* is neither germ cell-specific nor spermatid-specific by RT-PCR, RNA FISH reveals high reactivation levels in round spermatids. For spermatogonial RNA FISH quantitation (n=100 for each gene), cells were substaged based on characteristic cytological criteria. *Birc4*, *Fmr1* and *Zfx* are expressed in all spermatogonial subtypes, with 97%, 87% and 91% of combined

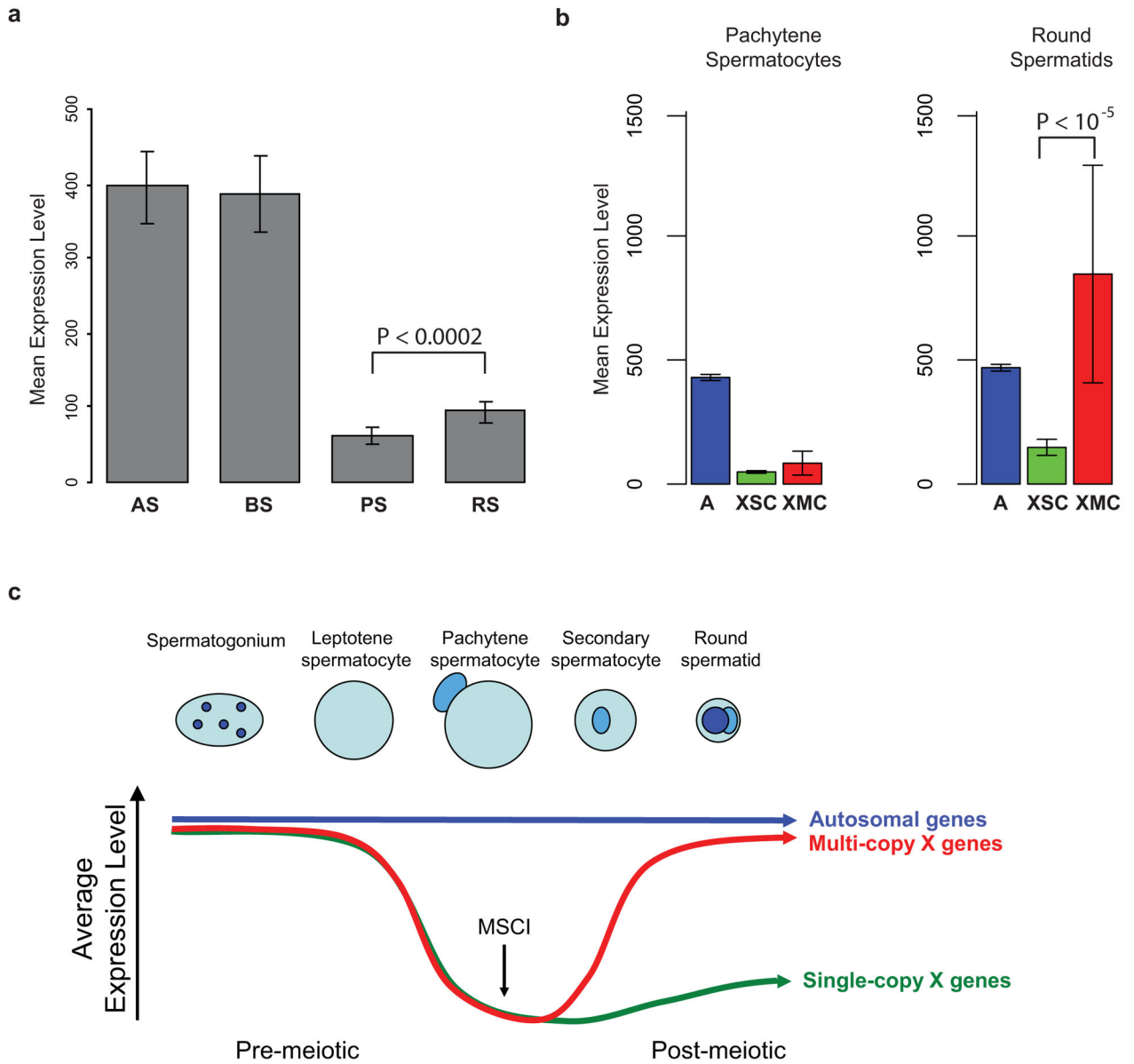
spermatogonia expressing, respectively. *Scml2* is expressed in 97% of type A spermatogonia but is silent in other spermatogonial substages. *Ott* is expressed in 99% of type A spermatogonia and 56% of intermediate and type B spermatogonia.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 5.**

Microarray analyses of single-copy and multi-copy genes on the X chromosome. **a**, Mean expression levels (+/- two standard error from mean) of 278 single-copy genes, previously defined as being expressed solely in spermatogonia7, in A and B spermatogonia (AS and BS), pachytene spermatocytes (PS) and round spermatids (RS). **b**, Mean expression levels of autosomal (A), X-linked single-copy (XSC), and X-linked multi-copy (XMC) genes in pachytene spermatocytes and round spermatids. P-values were determined via Wilcoxon Ranks Sum tests. **c**, Model: Multi-copy genes evade the effects of X chromosome post-meiotic repression. Single-copy and multi-copy X-linked genes exhibit similar average levels of expression during pre-meiotic spermatogenesis. All X-linked genes are subsequently silenced during MSCI. Following MSCI, single-copy X-linked genes exhibit

low reactivation levels, fitting with the post-meiotic repression of the X chromosome. Multi-copy X-linked genes exhibit expression levels similar to autosomal genes, thus evading the repressive effects of post-meiotic repression.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**

Ampliconic regions of the mouse X chromosome

	Map position (Mb)	Repeat unit size in kb	% Identity	Total ampliconic sequence in kb	Gene families	Copy number <sup>‡</sup>
Amp1	3.00–4.85	40.3 (105.0) 29.8 (108.0)	99.20 99.52	3,960 <sup>†</sup>	<i>Sstx</i> <i>Mgelh</i>	~8 ~21
Amp2	7.45–8.05	15.2	96.23	197.6	<i>Ssxb</i>	13
Amp3	8.10–8.25	23.4 (49.2)	93.23	98.4	<i>Fthl17</i> <i>Zfp-161-like</i>	6 2
Amp4	23.20–29.80	34.3 25.7	98.9 96.86	4,720 <sup>†</sup>	<i>Slx</i> <i>E330016L19Rik</i>	~25 ~21
Amp5	33.65–34.40	21.0	95.32	357.0	<i>Rhox</i>	28
Amp6	49.77–49.89	8.5 (42.3)	96.31	84.6	<i>Cxx1a/b/c</i>	3
Amp7	50.32–52.18	93.0 *	99.96*	2,400 <sup>†</sup>	<i>Slx-like 1</i>	~14
Amp8	66.65–66.81	25.2 (64.6)	99.59	129.2	<i>Magea</i> <i>4930567H17Rik</i>	2 2
Amp9	69.32–69.57	27.1 25.4 (38.6)	99.06 97.23	197.1	<i>Xlr</i>	12
Amp10	70.88–71.23	46.6 (53.7) 15.6 (19.8)	99.32 96.64	173.8	<i>Sstx</i> <i>Ube1C2-like</i>	3 2
Amp11	74.10–74.52	16.4 13.0	98.21 93.22	75.2	<i>Obpl</i>	7
Amp12	87.62–88.26	34.6 30.5 (39.8)	99.84 98.62	148.8	<i>Mageb</i>	8
Amp13	90.92–91.12	21.4 (52.0)	99.36	104.0	<i>ZxdB-like</i>	2
Amp14	98.86–99.17	17.4 (35.8) 19.8 (43.6)	98.12 95.31	294.4	<i>Dmrtc1b</i>	12
Amp15	99.18–99.31	13.0 (32.2)	99.03	64.4	<i>Pabp-like</i>	2
Amp16	114.53–114.61	17.1	95.68	34.2	<i>Tgifx</i>	2
Amp17	119.50–121.60	125.6*	99.89*	2,020 <sup>†</sup>	<i>Srsx</i> <i>LOC665542</i>	~14 ~12
Amp18	130.42–130.92	86.8	98.81	358.1	<i>Pramel-3</i>	9
Amp19	142.72–145.62	190.3*	97.74*	3,800 <sup>†</sup>	<i>Ott</i>	~12
Amp20	147.75–147.90	28.6	99.22	57.2	<i>LOC207318</i>	2
Amp21	149.33–149.48	17.4*	99.73*	69.6	<i>4921511MI7Rik</i>	4

	Map position (Mb)	Repeat unit size in kb	% Identity	Total ampliconic sequence in kb	Gene families	Copy number <sup>‡</sup>
Amp22	150.24–150.45	43.8*	99.17*	87.6	<i>Magea</i>	6
Total		1,474.5		19.43Mb		254

Map position coordinates are from the UCSC genome browser assembly from February 2006. Arm lengths for each repeat unit in an ampliconic region were defined by the IRF program. Seven ampliconic regions (Amp1, Amp4, Amp9, Amp10, Amp11, Amp12, and Amp14) contain two different repeat unit types. In some cases IRF-defined arms are truncated forms (due to indels and/or dissimilar sequence stretches, see Supplementary Fig. 1 for examples) of larger repeat units. Complete repeat unit sizes are shown in parenthesis adjacent to the IRF-defined sizes. Only genes landing entirely within a repeat arm are listed. Repeat units with > 90% arm-to-arm identity and arms with >8kb were selected from the IRF analysis. In cases when multiple repeat units can be paired, percent identities and repeat unit sizes are given for the largest repeat units. Genes with either testis-biased or testis-specific expression (Fig. 2) are shown in bold.

\* Tandem duplicated repeats not identified by IRF (percent identities and arm lengths were determined via dot-plots and ClustalW alignments).

<sup>‡</sup>Total amplicon sequence space estimates were based upon the number of BAC-end sequences (Supplementary Table 2).

<sup>‡</sup>Does not include pseudogenes identified in NCBI assembly 37.1.1. For unassembled ampliconic regions we provide estimates based only upon NCBI assembly 37.1.1, thus the gene copy number could change upon contiguous assembly of the region.