nature genetics



# The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression

Jacob L Mueller[1], Shantha K Mahadevaiah[2], Peter J Park[3], Peter E Warburton[4], David C Page[1] & James M A Turner[2]

**According to the prevailing view, mammalian X chromosomes are enriched in spermatogenesis genes expressed before meiosis[1–3] and deficient in spermatogenesis genes expressed after meiosis[2,3]. The paucity of postmeiotic 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[4,5]. Recent studies have concluded that MSCI-initiated silencing persists beyond meiosis[6–8] and that most genes on the X chromosome remain repressed in round spermatids[7]. Here, we report that 33 multicopy gene families, representing ~273 mouse X-linked genes, are expressed in the testis and that this expression is predominantly in postmeiotic cells. RNA FISH and microarray analysis show that the maintenance of X chromosome postmeiotic repression is incomplete. Furthermore, X-linked multicopy genes exhibit a similar degree of expression as 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 are expressed in postmeiotic cells.**

The existing model that the X chromosome is deficient in postmeiotic spermatogenesis genes was based primarily on the analysis of single-copy genes[2,3]. We therefore sought to explore the spermatogenesis expression profiles of X-linked multicopy genes. We carried out a systematic search for ampliconic regions—comprising palindromic or tandem segmental duplications—and their associated genes on the mouse X chromosome. To identify palindromic repeats, we used the inverted repeats finder (IRF) program[9], focusing on palindromes with large arms (≥8 kb) that showed ≥90% nucleotide identity between arms and that were <500 kb 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 multicopy gene clusters (**Supplementary Table 1** online). We found 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 totaled 1,474 kb (**Table 1**). Together, these ampliconic sequences comprise approximately 19.4 Mb, or ~12% of the ~166 Mb mouse X chromosome (**Table 1**).

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

We next asked whether mouse X-ampliconic genes show testis-biased expression. We searched X-ampliconic regions for ESTs and previously unknown or predicted genes and identified 26 multicopy genes located entirely within these ampliconic regions (**Supplementary Table 3** online). Through RT-PCR, we found that 23 of the 26 X-ampliconic multicopy genes are expressed predominantly or exclusively in the testis (**Fig. 2**). The 23 testis-expressed multicopy genes reside in 20 of the 22 ampliconic regions (**Fig. 1**). We examined the genomic structure of these 20 regions (**Supplementary Fig. 1** online) and found that the X-ampliconic gene copy number ranges from 2 to ~28 (**Fig. 1**), with a total of ~232 ampliconic protein-coding gene copies showing testis-biased expression.

To determine whether testis-biased expression of multicopy genes is a phenomenon limited to ampliconic regions, we also analyzed nonampliconic multicopy genes. We defined nonampliconic sequence as all regions of the mouse X chromosome not harboring ampliconic sequences (~88% of the mouse X chromosome). We selected genes present in at least two copies located within 1 Mb of each other and with ≥85% amino acid identity (**Supplementary Table 1**). We

[1]Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Division of Stem Cell Research and Developmental Genetics, Medical Research Council, National Institute for Medical Research, London NW7 1AA, UK. [3]Children's Hospital Informatics Program, Harvard-Partners Center for Genetics and Genomics, Boston, Massachusetts 02115, USA. [4]Department of Human Genetics, Mount Sinai School of Medicine, New York, New York 10029, USA. Correspondence should be addressed to J.M.A.T. (jturner@nimr.mrc.ac.uk) or J.L.M. (jmueller@wi.mit.edu).

Received 1 November 2007; accepted 15 February 2008; published online 4 May 2008; doi:10.1038/ng.126

**Table 1 Ampliconic regions of the mouse X chromosome**

| | Map position (Mb) | Repeat unit size in kb | Percentage identity | Total ampliconic sequence in kb | Gene families | Copy number[a] |
|---|---|---|---|---|---|---|
| Amp1 | 3.00–4.85 | 40.3 (105.0) | 99.20 | 3,960[b] | ***EG668965*** | ~8 |
| | | 29.8 (108.0) | 99.52 | | ***Gmcl1l*** | ~21 |
| Amp2 | 7.45–8.05 | 15.2 | 96.23 | 197.6 | ***Ssxb*** | 13 |
| Amp3 | 8.10–8.25 | 23.4 (49.2) | 93.23 | 98.4 | ***Fthl17*** | 6 |
| | | | | | ***Zfp161*** | 2 |
| Amp4 | 23.20–29.80 | 34.3 | 98.9 | 4,720[b] | ***Slx*** | ~25 |
| | | 25.7 | 96.86 | | ***E330016L19Rik*** | ~21 |
| Amp5 | 33.65–34.40 | 21.0 | 95.32 | 357.0 | ***Rhox*** | 28 |
| Amp6 | 49.77–49.89 | 8.5 (42.3) | 96.31 | 84.6 | *Cxx1* | 3 |
| Amp7 | 50.32–52.18 | 93.0[c] | 99.96[c] | 2,400[b] | ***4930527E24Rik*** | ~14 |
| Amp8 | 66.65–66.81 | 25.2 (64.6) | 99.59 | 129.2 | ***Magea*** | 2 |
| | | | | | ***4930567H17Rik*** | 2 |
| Amp9 | 69.32–69.57 | 27.1 | 99.06 | 197.1 | *Xlr* | 12 |
| | | 25.4 (38.6) | 97.23 | | | |
| Amp10 | 70.88–71.23 | 46.6 (53.7) | 99.32 | 173.8 | ***Sstx*** | 3 |
| | | 15.6 (19.8) | 96.64 | | ***EG434797*** | 2 |
| Amp11 | 74.10–74.52 | 16.4 | 98.21 | 75.2 | *Obp1* | 7 |
| | | 13.0 | 93.22 | | | |
| Amp12 | 87.62–88.26 | 34.6 | 99.84 | 148.8 | ***Mageb*** | 8 |
| | | 30.5 (39.8) | 98.62 | | | |
| Amp13 | 90.92–91.12 | 21.4 (52.0) | 99.36 | 104.0 | ***Zxd*** | 2 |
| Amp14 | 98.86–99.17 | 17.4 (35.8) | 98.12 | 294.4 | ***Dmrtc1b*** | 12 |
| | | 19.8 (43.6) | 95.31 | | | |
| Amp15 | 99.18–99.31 | 13.0 (32.2) | 99.03 | 64.4 | ***Pabpc1l2*** | 2 |
| Amp16 | 114.53–114.61 | 17.1 | 95.68 | 34.2 | ***Tgif2lx*** | 2 |
| Amp17 | 119.50–121.60 | 125.6[c] | 99.89[c] | 2,020[b] | ***Srsx*** | ~14 |
| | | | | | ***LOC665542*** | ~12 |
| Amp18 | 130.42–130.92 | 86.8 | 98.81 | 358.1 | ***Pramel3*** | 9 |
| Amp19 | 142.72–145.62 | 190.3[c] | 97.74[c] | 3,800[b] | ***Ott*** | ~12 |
| Amp20 | 147.75–147.90 | 28.6 | 99.22 | 57.2 | ***LOC207318*** | 2 |
| Amp21 | 149.33–149.48 | 17.4[c] | 99.73[c] | 69.6 | ***4921511M17Rik*** | 4 |
| Amp22 | 150.24–150.45 | 43.8[c] | 99.17[c] | 87.6 | ***Magea*** | 6 |
| Total | | 1,474.5 | | 19.43 Mb | | 254 |

Map position coordinates are from the University of California Santa Cruz Genome Browser (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 (as a result of indels and/or dissimilar sequence stretches, see **Supplementary Fig. 1** for examples) of larger repeat units. Complete repeat unit sizes are shown in parentheses 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 >8 kb were selected from the IRF analysis. In cases in which 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.

[a]Does not include pseudogenes identified in NCBI assembly 37.1. For unassembled ampliconic regions, we provide estimates based only upon NCBI assembly 37.1, thus the gene copy number could change upon contiguous assembly of the region. [b]Total amplicon sequence space estimates were based upon the number of BAC-end sequences (**Supplementary Table 2**). [c]Tandem duplicated repeats not identified by IRF (percent identities and arm lengths were determined via dot-plots and ClustalW alignments).

identified ten multicopy genes (**Fig. 1**), which ranged in copy number from 2 to 14, totaling to 41 gene copies. Like the X-ampliconic genes, all ten nonampliconic multicopy genes showed testis-biased expression (**Fig. 2**). This indicates that the mouse X chromosome has at least 33 multicopy genes (23 ampliconic and 10 nonampliconic) with testis-biased gene expression. Given that there are 1,555 X-linked genes (NCBI Build 37.1), then ~18% (232 ampliconic + 41 nonampliconic = 273 gene copies) of X-linked genes are multicopy genes with testis-biased expression, assuming that all copies of each family are expressed. Testis ESTs for multicopy family members (**Supplementary Table 3**) and copy-specific RT-PCR assays (**Supplementary Fig. 2** online) both suggested that most, if not all, of the 33 multicopy genes are expressed from more than a single copy.

We then determined whether the 33 multicopy X-linked genes are expressed in germ cells or somatic testis cells. We carried out RT-PCR on XX*Sry* and XX*Sxr*[b] testes that lack germ cells as a result of a combination of the absence of spermatogonial proliferation factor *Eif2s3y* as well as the toxic effects of double X chromosome dosage[11,12]. For 28 of the 33 multicopy genes, no expression was detected in either XX*Sry* or XX*Sxr*[b] testes (**Fig. 3a**), indicating that these 28 multicopy 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 multicopy X-linked genes initiate expression at or after 18.5 days post partum (d.p.p.) (**Fig. 3b**), when secondary spermatocytes and round spermatids first appear. The other eight X-linked genes initiate expression before meiosis, at 7.5 d.p.p. (**Fig. 3b**), and this early expression would mask any potential postmeiotic reactivation (**Fig. 4**). Despite this, we found that most (20 of 28) of X-linked multicopy genes are specifically expressed in postmeiotic spermatogenic cells. This finding is unexpected, as the X chromosome retains a repressed transcriptional state in postmeiotic cells[6–8].

To address the degree of postmeiotic repression, we examined nascent transcription of individual X-linked genes in round spermatids using RNA FISH (**Fig. 4**). We selected four single-copy X-linked genes: *Xiap*, *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[7]. 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** online), 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 carried out RNA FISH on mice carrying a ubiquitously expressed autosomal transgene (**Supplementary Fig. 3b**). All four X-linked genes were expressed in spermatogonia and then silenced in all late pachytene spermatocytes examined (**Fig. 4**; see legend for quantitation). This repression was maintained in most of the X-bearing round spermatids, but in contrast to previous reports, three of the four genes also showed some degree of reactivation: 0%, 7%, 17% and 18% for *Xiap, Scml2, Fmr1* and *Zfx*, respectively (see legend of **Fig. 4** and **Supplementary Table 4** online for details of quantification).

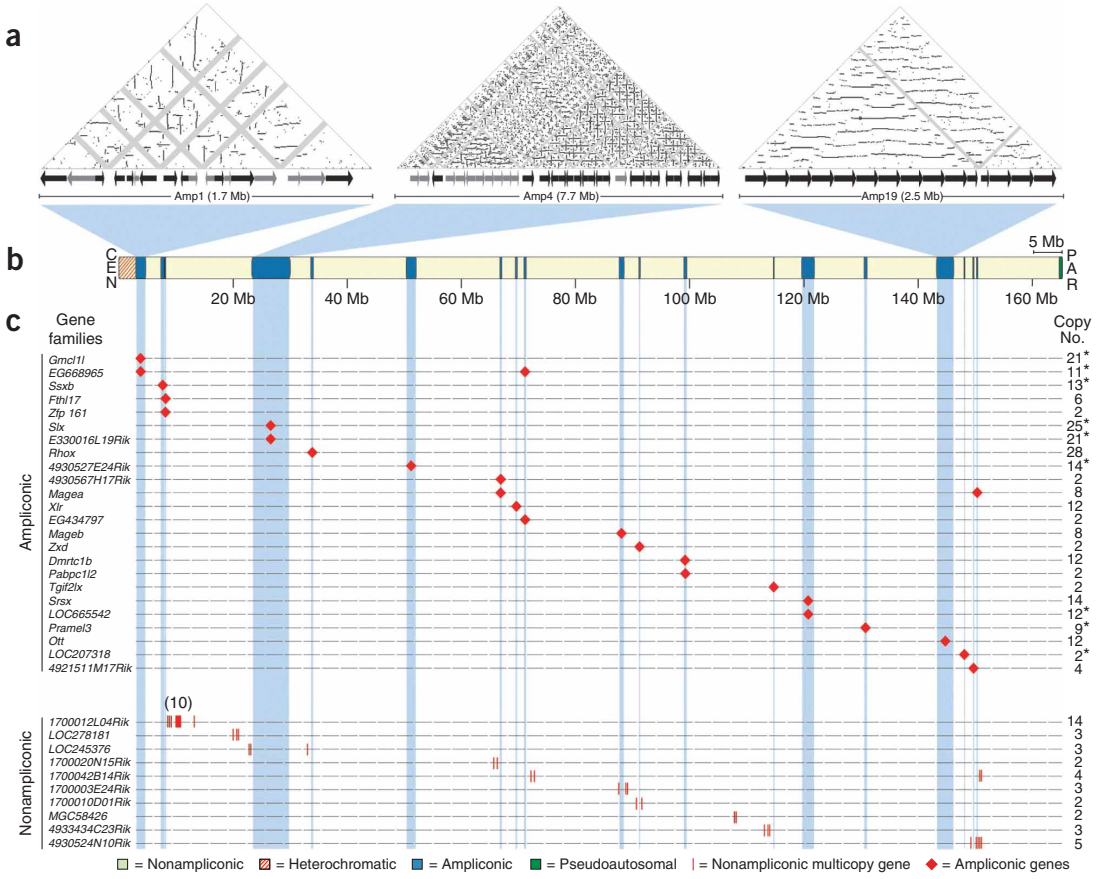


**Figure 1** Mouse X chromosome ampliconic regions containing testis-expressed genes. (**a**) Examples of complexity (Amp1, left), massive scale (Amp4, center) and tandem duplications (Amp19, right) in ampliconic regions. Each ampliconic region is compared to itself in a triangular dot-plot. Individual dots represent a perfect match of 200 nucleotides. Horizontal lines and vertical lines depict direct and inverted repeats, respectively. Shaded gray regions represent physical gaps in the sequence assembly. Primary (black) and secondary (gray) 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 nonampliconic multicopy genes and their copy number. Asterisks denote copy number estimates, as they fall within incompletely assembled genomic regions. Copy number estimates are based upon the number of intact ORFs in NCBI Build 37.1.

Do other single-copy X-linked genes behave the same way? We analyzed published microarray data from isolated germ cell populations to examine the postmeiotic expression of 278 single-copy genes classified as being expressed specifically in A or B spermatogonia[7]. We calculated the average expression values for these 278 genes in pachytene spermatocytes and round spermatids and found that the average expression in round spermatids is significantly higher than that in pachytene spermatocytes ($P < 0.0002$, Wilcoxon ranks-sum test, **Fig. 5a**). This suggests there is a general low level of postmeiotic 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, postmeiotic repression of the X chromosome is incomplete.

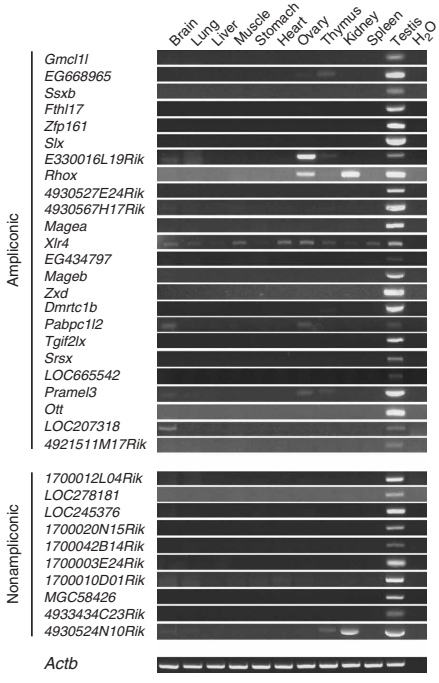


**Figure 2** Mouse X chromosome ampliconic and nonampliconic multicopy genes show testis-biased expression, as shown by RT-PCR. We assayed 26 ampliconic and 10 nonampliconic multicopy genes across 11 different tissues. *Cxx* and *Obp1* were not detected in any tissues and are therefore not shown. Of the 34 multicopy genes expressed in testis, only *Xlr4* did not show testis-biased expression. Other *Xlr* gene family members (*Xlr3* and *Xlr5*) show ubiquitous expression patterns similar to *Xlr4* (data not shown). *Actb*is ubiquitously expressed and serves as a control.

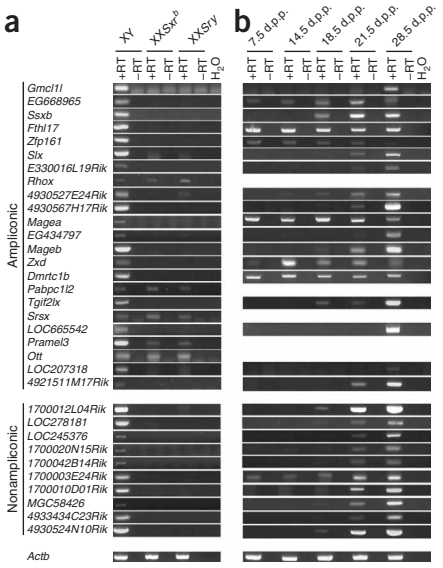


**Figure 3** Mouse X chromosome ampliconic and nonampliconic multicopy genes are expressed predominantly in germ cells during postmeiotic spermatogenesis, as shown by RT-PCR. (**a**) All 23 multicopy genes showing testis-biased expression were assayed in wild-type and germ cell–negative XX*Sry* and XX*Sxr*$^{b}$ testes. Only *Rhox*, *Pabpc1l2*, *Srsx*, *Pramel3* and *Ott* show expression in both XX*Sry* and XX*Sxr*$^{b}$ testes. We note that *Pramel3* gives faint products, consistent with previous findings[1] that the predominant site of expression for this multicopy gene is in germ cells. (**b**) Genes showing strict germ cell–specificity were assayed by analysis of the first wave of spermatogenesis in juvenile male mice. 20 of the 28 genes are expressed from 18.5 d.p.p., and the remaining 8—*EG668965, Fthl17, Zfp161, Magea, Zxd, Dmrtc1b, 1700003E24Rik* and *1700010D01Rik*—are expressed from 7.5 d.p.p.

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 multicopy gene. To test this, we examined three multicopy genes by RNA FISH: *Ott* (~12 copies), *4930527E24Rik* (~14 copies) and *Slx* (~25 copies). Unlike the single-copy X-linked genes, all three multicopy genes showed RNA FISH signals in most of the X-bearing round spermatids (**Fig. 4**). Furthermore, the percentage of X-bearing round spermatids that showed expression increased with gene copy number: 54% for *Ott*, 78% for *4930527E24Rik* and 93% for *Slx*. For comparison, we examined the frequency of round spermatid expression for a single-copy spermatid-specific autosomal gene, *Zscan2*. We detected *Zscan2* RNA FISH signals in 64% of round spermatids, a frequency similar to that of the three X-linked multicopy genes. The

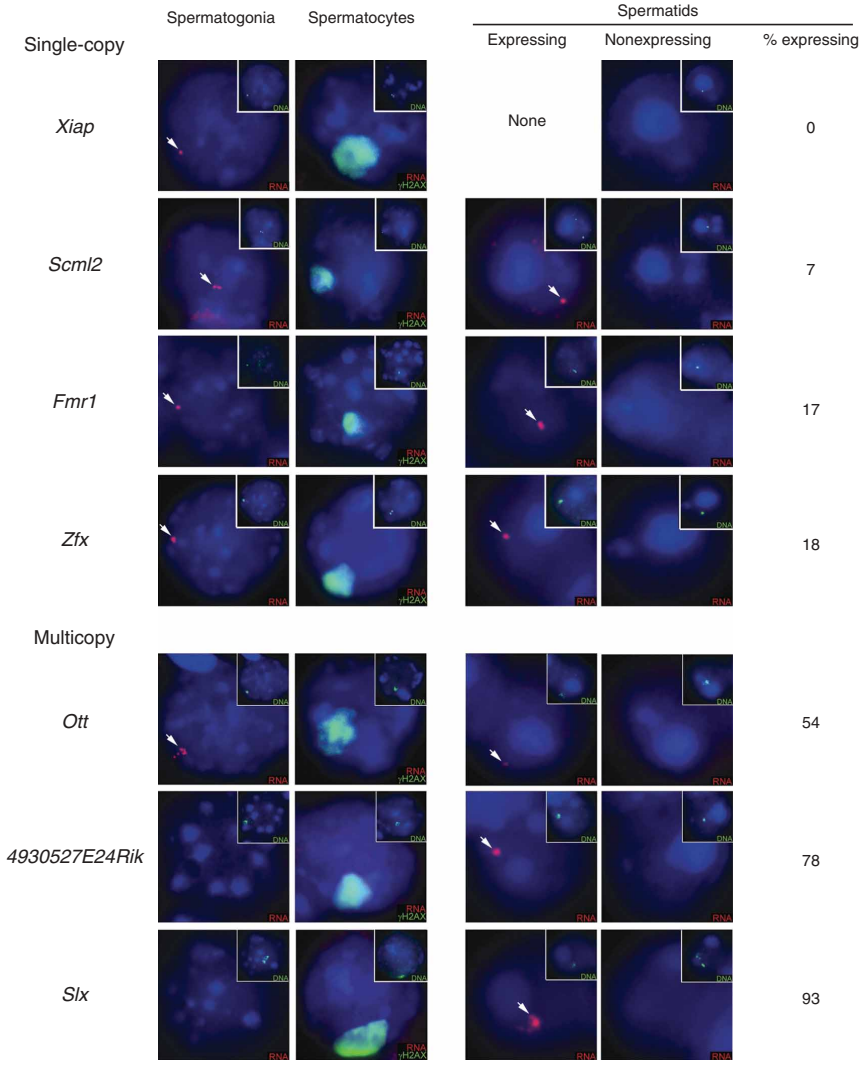


**Figure 4** Multicopy ampliconic genes show a greater degree of reactivation in round spermatids than single-copy genes. Four single-copy and three multicopy X-linked genes were assayed for expression during successive stages of spermatogenesis by RNA/DNA FISH. First column, spermatogonia. All four single-copy genes are expressed at this stage, consistent with previously reported observations (see below for details of quantitation)[7]. In contrast, only one multicopy gene, *Ott*, is expressed. Second column, pachytene spermatocytes. The sex body is shown by γH2AX 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**). *Xiap* expression could not be detected in any round spermatids, and so the 'spermatid-expressing' column is blank for this gene. In contrast, multicopy 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 results indicate high reactivation levels in round spermatids. For spermatogonial RNA FISH quantitation ($n = 100$ for each gene), cells were substaged on the basis of characteristic cytological criteria. *Xiap*, *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.

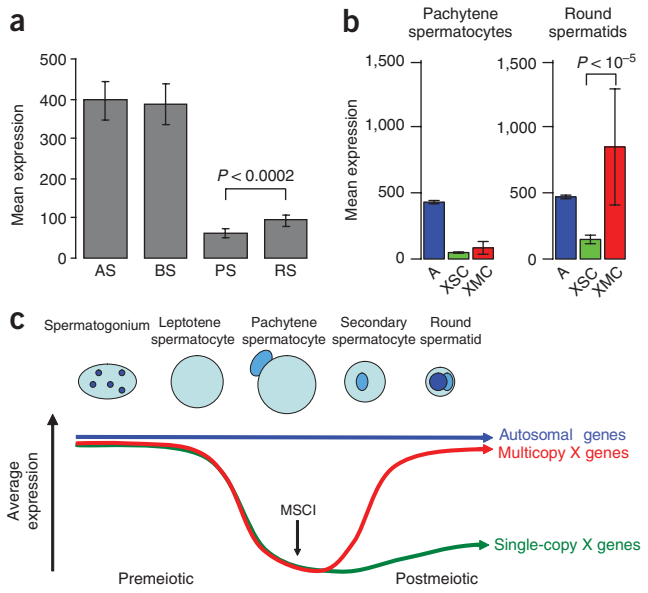


**Figure 5** Microarray analyses of single-copy and multicopy genes on the X chromosome. (**a**) Mean expression (± 2 s.e.m.) of 278 single-copy genes, previously defined as being expressed solely in spermatogonia[7], in A and B spermatogonia (AS and BS), pachytene spermatocytes (PS) and round spermatids (RS). (**b**) Mean expression 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**) Multicopy genes evade the effects of X chromosome postmeiotic repression. Single-copy and multicopy X-linked genes show similar average expression during premeiotic spermatogenesis. All X-linked genes are subsequently silenced during MSCI. Following MSCI, single-copy X-linked genes show low reactivation levels, fitting with the postmeiotic repression of the X chromosome. Multicopy X-linked genes show expression levels similar to autosomal genes, thus evading the repressive effects of postmeiotic repression.

high percentage of X-bearing round spermatids expressing a given multicopy gene—in contrast to the low percentage expressing single-copy X-linked genes—suggests that gene amplification facilitates higher gene expression in the face of postmeiotic X chromosome repression.

Do multicopy X-linked genes show higher spermatid expression 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[7]. We identified 24 X-linked multicopy genes, 552 X-linked single-copy genes and 20,236 autosomal genes. In pachytene spermatocytes, both X-linked multicopy and single-copy genes show expression consistent with the presence of MSCI (**Fig. 5b**). However, in round spermatids, X-linked multicopy genes show a significantly higher average expression than X-linked single-copy genes ($P < 10^{-5}$, Wilcoxon ranks-sum test, **Fig. 5b** and **Supplementary Table 5** online). This high degree of postmeiotic X-linked multicopy gene expression further suggests that the amplification of genes on the X chromosome compensates for postmeiotic repression.

Evolutionary models predict the accumulation of male advantage alleles on X chromosomes[13]. X-linked spermatogenesis genes identified before this study are primarily expressed in premeiotic cells[1,2]. Here we describe a collection of 33 X-linked multicopy genes that are expressed in postmeiotic spermatogenic cells. Some of these gene families may have important roles during sperm maturation, including *Gmcl1l*, whose autosomal paralog *Mgcl1* is required for acrosome formation and spermatid chromatin condensation[14], and *1700012L04Rik*, which encodes a recently identified spermatid-specific histone variant, H2AL1 (ref. 15).

To understand the unusual expression pattern of X-linked multicopy genes in the context of postmeiotic repression, we have re-examined the extent of X-chromosome repression during spermiogenesis. Postmeiotic X repression is incomplete, as single-copy X-linked genes with premeiotic expression show varying extents of reactivation. This reconciles the apparent conflict between studies documenting the existence of a repressed X chromosome in postmeiotic spermatogenic cells[6–8] and those showing reactivation of selected single-copy X-linked genes[16]. Notably, we find that X-linked multicopy genes yield an elevated frequency of expressing X-bearing round spermatids and that the average expression 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 postmeiotic repression, because rare cases of robust reactivation of single-copy X-linked genes (for example, *Uba1*) have also been identified[8]. Nevertheless, amplification of X-linked genes may have evolved to compensate for the repressive chromatin environment affecting the X chromosome in postmeiotic cells (**Fig. 5c**). Indeed, the human X chromosome harbors multicopy genes similar to those described here as well as primate-specific multicopy genes showing postmeiotic expression[9,17,18].

## METHODS

**Identification of amplicons and multicopy genes.** Individual mouse X chromosome amplicons (palindromic or tandem segmental duplications) were first identified using the inverted repeats finder (IRF) program[9]. We used the *Mus musculus* reference sequence (NCBI Build 36.1) to select palindromes that had arms ≥8 kb and ≥90% nucleotide identity between arms, and that were <500 kb of each other. The ≥8 kb arm size restriction, used in previous palindrome analyses[9], should eliminate all repeats due to recent LINE insertions. Custom tracks can be downloaded (see URLs section below) to display all mouse X chromosome palindromes on the University of California Santa Cruz 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, we used custom perl code[19] to characterize their genomic structure via triangular dot-plots (**Supplementary Fig. 1**).

As the IRF program identifies only palindromes, we used an alternative strategy to detect amplicons in tandem, which could also confirm the IRF-identified amplicons. We used the Biomart program (Ensembl) to obtain all X-linked protein-coding genes with paralogous genes on the X chromosome (**Supplementary Table 1**). Multicopy genes less than 1 Mb apart, which share ≥80% amino acid identity between any two copies, and which are not found in multiple copies throughout the mouse genome (for example, retroviral proteins, ribosomal proteins, or olfactory receptors), were considered putative ampliconic regions. Multicopy gene clusters passing these three criteria (**Supplementary Table 1**) were subjected to triangular dot-plot analysis[19] to determine whether the region was ampliconic. Ampliconic regions identified by 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[20]. Multicopy gene clusters fulfilling the three selection criteria, but not present in ampliconic regions were termed nonampliconic multicopy 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 (see URLs section below) and subsequently compared to the NCBI mouse EST database using BLAST[21]. 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 newly identified and known genes identified via BLAST searches.

**RT-PCR and RNA/DNA FISH.** We extracted total RNA using Trizol (Gibco BRL) according to manufacturer's instructions. For DNase treatment, 3 μg total RNA was incubated with 1.5 U RQ DNase 1 (Promega) at 37 °C for 90 min. The RNA was precipitated, resuspended and then reverse transcribed using Superscript II reverse transcriptase (200 U, Gibco BRL) for 90 min at 42 °C. We carried out PCR using the parameters: 1 cycle at 94 °C (3 min), 35 cycles at 94 °C (30 s), 56 °C (30 s), 72 °C (30 s) and then 1 cycle at 72 °C (10 min). 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 previously described[8]. We used long-range PCR products from BACs as probes for *Xiap, Fmr1, Scml2, Ott, 4930527E24Rik, Slx,* and *Zfx*. Primer sequences and BAC identifier names are listed in **Supplementary Table 3**.

**Microarray and statistical analyses.** We analyzed published microarray expression data (see Accession codes section below)[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 previously described[7] to set the trimmed (2% from each side) mean signal intensity to 125. Probes with $<100$ signal intensity at all time points 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 probe IDs and expression values for the 25 multicopy genes for each time point are listed in **Supplementary Table 5**. When multiple probes matched a multicopy gene (see **Supplementary Table 5** for genes with multiple probe_ids), we selected a single probe, with the median signal intensity. We carried out Wilcoxon ranks-sum paired tests using JMP 5.1 software (SAS) on all comparisons of average expression levels, as they diminish the influence of outliers.

**URLs.** Laboratory of Biocomputing and Informatics, https://tandem.bu.edu; Repeat Masker, http://www.repeatmasker.org.

**Accession codes.** NCBI GEO: microarray data, GSE4193.

*Note: Supplementary information is available on the Nature Genetics website.*

**ACKNOWLEDGMENTS**

We thank J. Alfoldi, D. Bellott, P. Burgoyne, H. Byers, J. Hughes, J. Lange, L. Reynard, S. Rozen and H. Skaletsky for helpful comments on the manuscript; D. Bellott, M. Gill, Y. Hu, J. Hughes and H. Skaletsky for technical advice; and P. Burgoyne and M. Mitchell for the *Uty* transgenic line. This work was supported by the Medical Research Council, US National Institutes of Health grant HD000257 (to D.C.P.), including fellowship F32HD052379 (to J.L.M.), and the Howard Hughes Medical Institute.

**AUTHOR CONTRIBUTIONS**

J.L.M., J.M.A.T., S.K.M. and P.E.W. identified amplicons and multicopy genes. J.L.M., J.M.A.T. and S.K.M. carried out RT-PCRs and RNA FISH. J.L.M. and P.J.P. did microarray and statistical analysis. The paper was written by J.L.M., J.M.A.T. and D.C.P.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Wang, P.J., McCarrey, J.R., Yang, F. & Page, D.C. An abundance of X-linked genes expressed in spermatogonia. *Nat. Genet.* **27**, 422–426 (2001).
2. Khil, P.P., Smirnova, N.A., Romanienko, P.J. & Camerini-Otero, R.D. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat. Genet.* **36**, 642–646 (2004).
3. Reinke, V. Sex and the genome. *Nat. Genet.* **36**, 548–549 (2004).
4. McKee, B.D. & Handel, M.A. Sex chromosomes, recombination, and chromatin conformation. *Chromosoma* **102**, 71–80 (1993).
5. Turner, J.M. Meiotic sex chromosome inactivation. *Development* **134**, 1823–1831 (2007).
6. Greaves, I.K., Rangasamy, D., Devoy, M., Marshall Graves, J.A. & Tremethick, D.J. The X and Y chromosomes assemble into H2A.Z-containing facultative heterochromatin following meiosis. *Mol. Cell. Biol.* **26**, 5394–5405 (2006).
7. Namekawa, S.H. *et al.* Postmeiotic sex chromatin in the male germline of mice. *Curr. Biol.* **16**, 660–667 (2006).
8. Turner, J.M., Mahadevaiah, S.K., Ellis, P.J., Mitchell, M.J. & Burgoyne, P.S. Pachytene asynapsis drives meiotic sex chromosome inactivation and leads to substantial post-meiotic repression in spermatids. *Dev. Cell* **10**, 521–529 (2006).
9. Warburton, P.E., 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.* **14**, 1861–1869 (2004).
10. Eichler, E.E., Clark, R.A. & She, X. An assessment of the sequence gaps: unfinished business in a finished human genome. *Nat. Rev. Genet.* **5**, 345–354 (2004).
11. Mroz, K., Carrel, L. & Hunt, P.A. Germ cell development in the XXY mouse: evidence that X chromosome reactivation is independent of sexual differentiation. *Dev. Biol.* **207**, 229–238 (1999).
12. Mazeyrat, S. *et al.* A Y-encoded subunit of the translation initiation factor Eif2 is essential for mouse spermatogenesis. *Nat. Genet.* **29**, 49–53 (2001).
13. Rice, W.R. Sex chromosomes and the evolution of sexual dimorphism. *Evolution Int. J. Org. Evolution* **38**, 735–742 (1984).
14. Kimura, T. *et al.* Mouse germ cell-less as an essential component for nuclear integrity. *Mol. Cell. Biol.* **23**, 1304–1315 (2003).
15. Govin, J. *et al.* Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J. Cell Biol.* **176**, 283–294 (2007).
16. Wang, P.J., Page, D.C. & McCarrey, J.R. Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse. *Hum. Mol. Genet.* **14**, 2911–2918 (2005).
17. Westbrook, V.A. *et al.* Spermatid-specific expression of the novel X-linked gene product SPAN-X localized to the nucleus of human spermatozoa. *Biol. Reprod.* **63**, 469–481 (2000).
18. Westbrook, V.A. *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.* **12**, 703–716 (2006).
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.* **29**, 279–286 (2001).
20. Waterston, R.H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
21. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).