doi:10.1038/nature13206

Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators

Daniel W. Bellott[1], Jennifer F. Hughes[1], Helen Skaletsky[1], Laura G. Brown[1], Tatyana Pyntikova[1], Ting-Jan Cho[1], Natalia Koutseva[1], Sara Zaghlul[1], Tina Graves[2], Susie Rock[2], Colin Kremitzki[2], Robert S. Fulton[2], Shannon Dugan[3], Yan Ding[3], Donna Morton[3], Ziad Khan[3], Lora Lewis[3], Christian Buhay[3], Qiaoyan Wang[3], Jennifer Watt[3], Michael Holder[3], Sandy Lee[3], Lynne Nazareth[3], Steve Rozen[1], Donna M. Muzny[3], Wesley C. Warren[2], Richard A. Gibbs[3], Richard K. Wilson[2] & David C. Page[1]

The human X and Y chromosomes evolved from an ordinary pair of autosomes, but millions of years ago genetic decay ravaged the Y chromosome, and only three per cent of its ancestral genes survived. We reconstructed the evolution of the Y chromosome across eight mammals to identify biases in gene content and the selective pressures that preserved the surviving ancestral genes. Our findings indicate that survival was nonrandom, and in two cases, convergent across placental and marsupial mammals. We conclude that the gene content of the Y chromosome became specialized through selection to maintain the ancestral dosage of homologous X–Y gene pairs that function as broadly expressed regulators of transcription, translation and protein stability. We propose that beyond its roles in testis determination and spermatogenesis, the Y chromosome is essential for male viability, and has unappreciated roles in Turner's syndrome and in phenotypic differences between the sexes in health and disease.

The human X and Y chromosomes evolved from autosomes over the past 300 million years[1]. Only 3% of ancestral genes survive on the human Y chromosome[2,3], compared to 98% on the X chromosome[4]. Y-chromosome decay was initially rapid but has virtually halted over the last 25 million years, leaving a stable set of ancestral genes[5–7]. Mathematical models of Y-chromosome decay assume all ancestral genes are equally likely to survive. However, our initial studies of the human Y chromosome suggested that its gene content is functionally coherent[8], leading us to ask whether mammalian Y chromosomes preferentially retained a subset of ancestral genes, and, if so, what qualities these surviving genes share.

Our earlier analyses[8] of the human Y chromosome were hampered by limited knowledge of the gene content of the ancestral autosomes. Our recent cross-species comparisons enabled us to reconstruct their gene content and identify acquired genes on the X and Y chromosomes. The human X chromosome acquired and amplified testis-expressed gene families[2,4]. Similarly, our comparisons of the human, chimpanzee and rhesus Y chromosomes indicated recent acquisition and amplification of testis-specific genes[3,5,6]. Thus, both the human X and Y chromosomes gained a specialization for male reproduction by acquiring genes that were not present on the ancestral autosomes[2–4].

We excluded acquired genes to independently examine ancestral Y-linked genes for characteristics that distinguished surviving genes from genes lost to decay. Because the human, chimpanzee and rhesus Y chromosomes share nearly identical ancestral gene content, we analysed five additional mammals to enhance our ability to detect biases in the decay and survival of ancestral genes. We produced finished sequence of the ancestral portions of the Y chromosomes of marmoset (*Callithrix jacchus*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), bull (*Bos taurus*) and opossum (*Monodelphis domestica*) and compared them to the published sequences of the human, chimpanzee (*Pan trogolodytes*) and rhesus macaque (*Macaca mulatta*) Y chromosomes, all eight corresponding X chromosomes and the orthologous chicken (*Gallus gallus*) autosomes as an outgroup to mammalian X and Y chromosomes. Using this expanded tree of species, we reconstructed the evolution of mammalian Y chromosomes from their origin to the present. We concluded that surviving Y-linked genes form a functionally coherent group enriched for dosage-sensitive, broadly expressed regulators of transcription, translation and protein stability.

We produced finished sequence using the SHIMS (single-haplotype iterative mapping and sequencing) strategy we employed on primate Y, human X and chicken Z chromosomes (Methods)[2–7]. These sequences comprise 17 megabases (Mb) and are accurate to about 1 nucleotide per 0.3 Mb (Supplementary Table 1, Extended Data Fig. 1 and Methods). To identify ancestral X–Y gene pairs, we searched for Y-homologues of protein-coding genes we had identified as ancestral (Supplementary Tables 2 and 3)[2,5]. We validated each putative gene by verifying transcriptional activity (Extended Data Fig. 2) and comparing its open reading frame to its chicken orthologue (Supplementary Data 1 and 2). We identified 36 different ancestral X–Y gene pairs across all eight species, adding 18 ancestral X–Y gene pairs to the 18 known to be present on the human, chimpanzee and rhesus Y chromosomes (Fig. 1).

Regulatory functions of X–Y gene pairs

Seventeen years ago, we characterized human X–Y gene pairs as specialized in cellular housekeeping functions[8]. Since then, annotation of the human genome has increased in detail and completeness. We therefore revisited the question of functional coherence and found evidence that X–Y pair genes perform an array of regulatory functions (Fig. 2). Based on annotations of their X homologues, ancestral Y-linked genes appear to regulate each stage of the central dogma: histone lysine demethylases *KDM5D* (H3K4) and *UTY* (H3K27); the transcription factor *ZFY*, regulating stem-cell self-renewal; spliceosomal component *RBMY*; translation initiation factors *DDX3Y* and *EIF1AY*; and the deubiquitinase *USP9Y* (Fig. 2). Compared to other ancestral genes that survive on the X chromosome, X–Y pair genes are enriched for annotations such as nucleic-acid binding, transcription and translation (Extended Data Table 1, Methods and Supplementary Table 4), suggesting that X–Y pair genes can govern expression of targets throughout the genome.

[1] Whitehead Institute, Howard Hughes Medical Institute, & Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [2] The Genome Institute, Washington University School of Medicine, St. Louis, Missouri 63108, USA. [3] Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA.

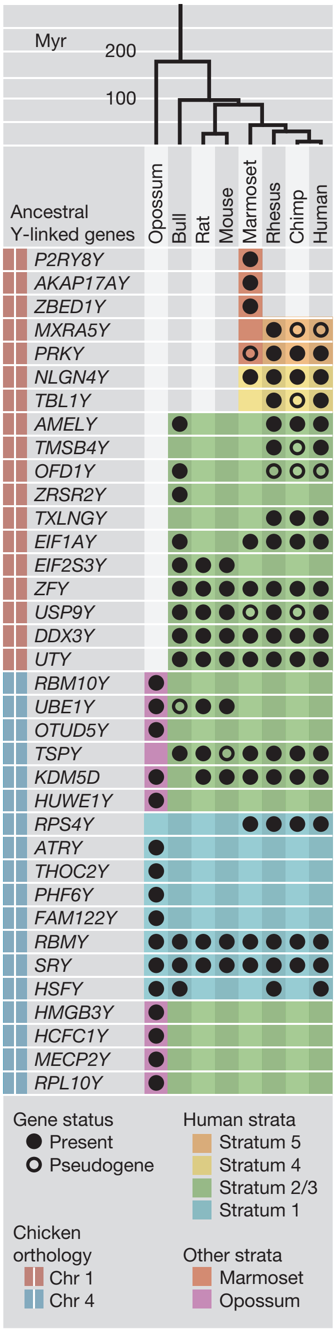


Figure 1 | Ancestral Y-linked genes by species and human X homologue location. Ancestral Y-linked genes (filled circles) and pseudogenes (open circles) listed by the position of their X-linked homologue on the human X chromosome. The placental-specific added region (red bar) and the conserved region shared with marsupials (blue bar) of the sex chromosomes are indicated on the left. Human sex chromosome evolution was punctuated by formation of at least 4 evolutionary strata (light blue, green, yellow and orange); other strata formed independently in opossum (purple) and marmoset (red). Myr, million years.

Convergent survival of X–Y gene pairs

To gain insight into the decay and survival of ancestral genes, we reconstructed Y chromosome evolution, taking advantage of our earlier discovery that synonymous nucleotide divergence between the X and Y sequences of each gene pair increases in stepwise fashion along the human X chromosome[1,3,9]. This suggested a series of discrete events, most likely inversions on the Y chromosome, that suppressed X–Y crossing over in a single region, or 'stratum,' without disturbing gene order on the X chromosome[1,9]. We used the 36 X–Y gene pairs to recalibrate previous reconstructions of evolutionary strata (Extended Data Table 2, Extended Data Figs 3–5, Methods and Supplementary Tables 2 and 5). In broad agreement with previous reconstructions[1–3,9,10], we concluded that the human X and Y chromosomes evolved from ordinary autosomes through chromosomal fusion and formation of at least four strata (Fig. 3 and Methods).

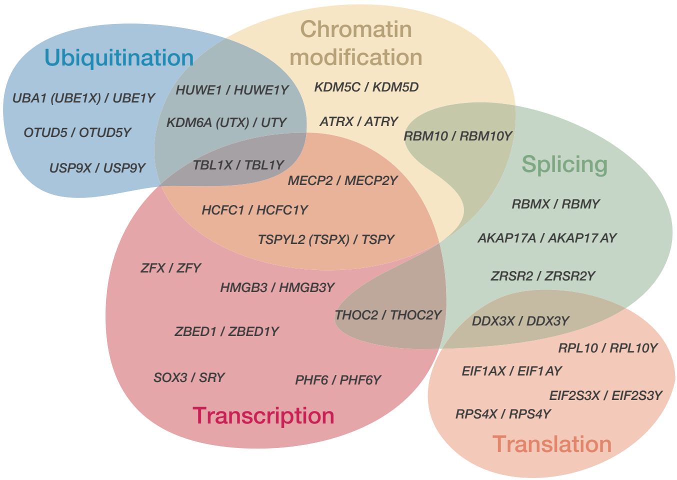


Figure 2 | Regulatory annotations of X–Y pair genes. Venn diagram depicting regulatory functions predicted for selected X–Y pair genes on basis of UniProt annotations of human X-homologue. Common alternatives to official gene symbols in parentheses.

Our results indicate that the stratum containing *UBE1Y* and *KDM5D* formed independently in the placental and marsupial lineages (Extended Data Fig. 4). The same set of ancestral genes became subject to genetic decay in each lineage, forming replicates of the same natural experiment. Out of the 184 ancestral genes shared between these strata, nine survived on the Y chromosome in marsupials, and three survived in placental mammals, but both lineages retained *UBE1Y* and *KDM5D* (Fig. 1, Supplementary Table 2 and Methods). The convergent survival of two ancestral genes is unlikely to occur under a model where genes survive genetic decay at random (one-tailed Fisher's exact test, $P < 6.25 \times 10^{-3}$).

Remarkable longevity of X–Y gene pairs

Using these recalibrated evolutionary strata, we re-examined the kinetics of genetic decay among ancestral Y-chromosome genes. Analysis of primate Y chromosomes had led us to conclude that, within a stratum, rapid gene loss was followed by stabilization at a baseline set of genes[5]. With five more divergent mammals, we doubled the constraints on the kinetics of gene loss during human Y chromosome evolution (Fig. 4 and Methods) and traced the stability of human Y-chromosome genes to the origin of mammals (Fig. 4). We infer that 97 million years ago, the Y chromosome of the common ancestor of placental mammals carried 18 ancestral genes from stratum 1 and stratum 2/3 (Fig. 1). Of those 18 genes, 14 survive in the human lineage (Fig. 1), and none have been lost in the last 44 million years (Fig. 4). We also examined whether ancestral Y-linked genes were stable in marsupials. Recent analyses of the tammar wallaby (*Macropus eugenii*) Y chromosome identified ten genes shared with the Tasmanian devil (*Sarcophilus harrisii*)[11]; we observe that all are ancestral and survive in the opossum. This suggests the opossum lineage maintained these genes over the last 78 million years[12]. We conclude that in both placental and marsupial lineages, some ancestral X–Y gene pairs were remarkably long lived despite rapid decay of surrounding genes.

Two strategies preserved Y-linked genes

In light of the regulatory annotations of X–Y gene pairs, convergent survival of X–Y gene pairs in the placental and marsupial lineages, and the longevity of ancestral X–Y gene pairs across mammals, we sought

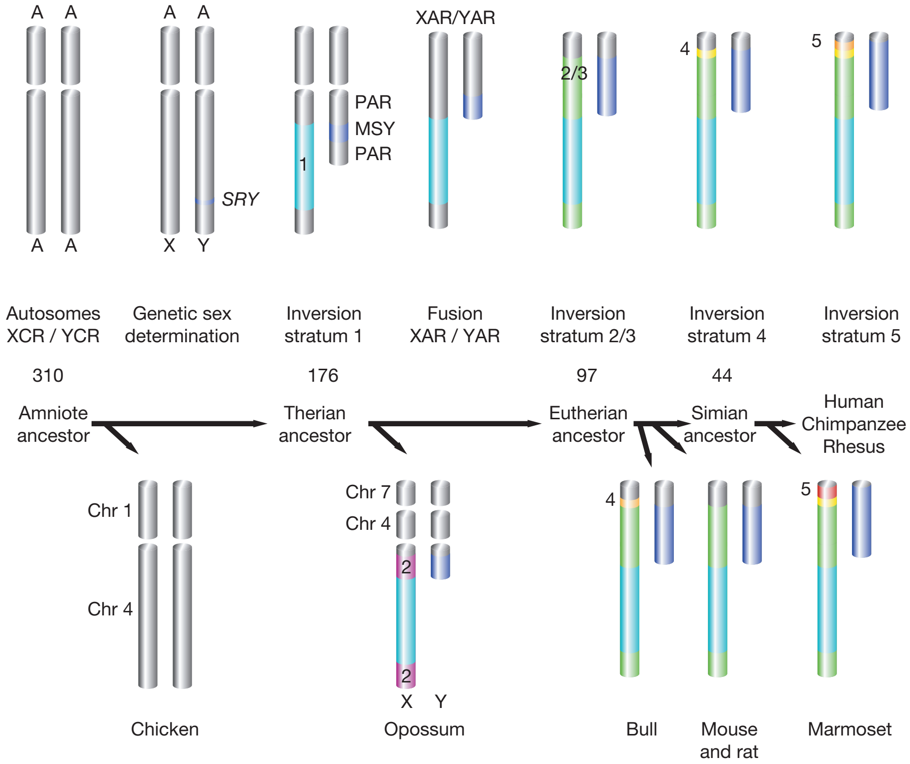


Figure 3 | Reconstruction of human sex chromosome evolution. Major events in the evolution of the human sex chromsomes are labelled with approximate dates in Myr. After *SRY* evolved, at least 4 evolutionary strata (light blue, green, yellow and orange) formed in the lineage leading to the human Y chromosome. Each stratum expanded the MSY (male-specific region of the Y, deep blue) at the expense of the PAR (pseudoautosomal region, grey). Genetic decay eliminated most genes from MSY. A chromosomal fusion extended the PAR, generating conserved (XCR/YCR) and added (XAR/YAR) regions.

the evolutionary pressures that drove their survival. We had previously speculated that biases in the gene content of the human Y chromosome could arise through two evolutionary strategies: retention and amplification of testis-specific gene families, and conservation of ancestral X–Y gene pairs to maintain comparable expression between males and females[8]. Using the set of 639 ancestral genes reconstructed through cross-species comparisons of the human X chromosome and orthologous chicken autosomes[2,4,5], we tested whether these hypotheses account for the 36 ancestral X–Y pair genes found on eight present-day Y chromosomes.

The Y chromosome was predicted to accumulate genes that enhance male reproductive fitness[13], which depends upon sperm production in the adult testis. In each species we studied, ancestral genes that are amplified into multi-copy families are expressed exclusively or predominantly in the testis (Extended Data Fig. 2). However, many such genes have broadly expressed single-copy homologues on orthologous chicken autosomes, on mammalian X chromosomes, and in cases like *DDX3Y*, *EIF1AY*, *UBE1Y* and *ZFY*, on other Y chromosomes (Extended Data Fig. 2 and Supplementary Table 2). This suggests that adoption of testis-specific function preceded gene amplification.

In light of evidence that intrachromosomal gene conversion preserved testis-specific gene families in primate Y-chromosome palindromes[14], we speculated that gene amplification contributed to longevity. We ranked surviving Y-linked genes by total branch length across our tree of eight species (Fig. 5a)[12]. Genes that are amplified in at least one species have a significantly greater branch length than those that are single copy in every species (one-tailed Mann–Whitney *U*-test, $P < 4.27 \times 10^{-5}$) (Fig. 5a). This correlation remains robust when the opossum lineage, with a large number of unique single-copy genes, is excluded (one-tailed Mann–Whitney *U*-test, $P < 5.54 \times 10^{-4}$). Gene families in tandem arrays show high intraspecies identity and interspecies divergence, a sign that gene conversion is more frequent than mutation in these structures (Extended Data Fig. 6). Two pairs of Y-linked genes, *RPS4Y1* and *RPS4Y2* in primates and *Zfy1* and *Zfy2* in mouse, are exceptions. Both are physically dispersed and show no sign of recent Y–Y gene conversion (Extended Data Fig. 6). We conclude that genes specialized for male reproduction avoided genetic decay through intrachromosomal gene conversion among members of a Y-linked, multi-copy gene family.

Next, we examined whether single-copy genes on the Y chromosome survived owing to selection to preserve the correct dosage of broadly expressed genes critical to both sexes[3,8,15]. Most genes on the Y chromosome were lost to genetic decay, and the X chromosome evolved mechanisms to compensate for the lost dosage of Y-linked genes in males[8,16,17]. The Y chromosome might preferentially retain genes for which the transition state of this process, with a non-functional Y-linked gene and a functional but non-dosage-compensated X-linked homologue, was disadvantageous. Dosage-sensitive genes functioning in many tissues and cell types might be particularly sensitive to these pressures[15]. We re-analysed published data sets for evidence that our set of 36 X–Y pair genes systematically differ from the 603 other ancestral genes on the X chromosome with regard to dosage sensitivity[18–20], breadth of expression[21,22] and intensity of purifying selection[23].

We examined whether X–Y pair genes show signs of dosage sensitivity. In humans, gene-by-gene estimates predict a greater likelihood of haploinsufficiency[18] for ancestral X-linked genes with surviving Y homologues compared to those lacking Y homologues (one-tailed Mann–Whitney *U*-test, $P < 6.59 \times 10^{-3}$) (Fig. 5b). If surviving X–Y gene pairs maintain ancestral gene dosage, then X-linked genes with surviving Y-linked homologues should escape X inactivation. In human[19], mouse[20], and opossum[24], data on allele-specific expression in females is informative for a subset of ancestral genes (Supplementary Table 2). In each species, a higher proportion of X-linked genes with surviving Y-linked homologues escape X-inactivation compared to those without surviving Y-linked homologues (Supplementary Table 2), and X–Y gene pairs in which the X-homologue is subject to X-inactivation have Y-homologues that show signs of functional differentiation. In humans, 12 of 14 informative

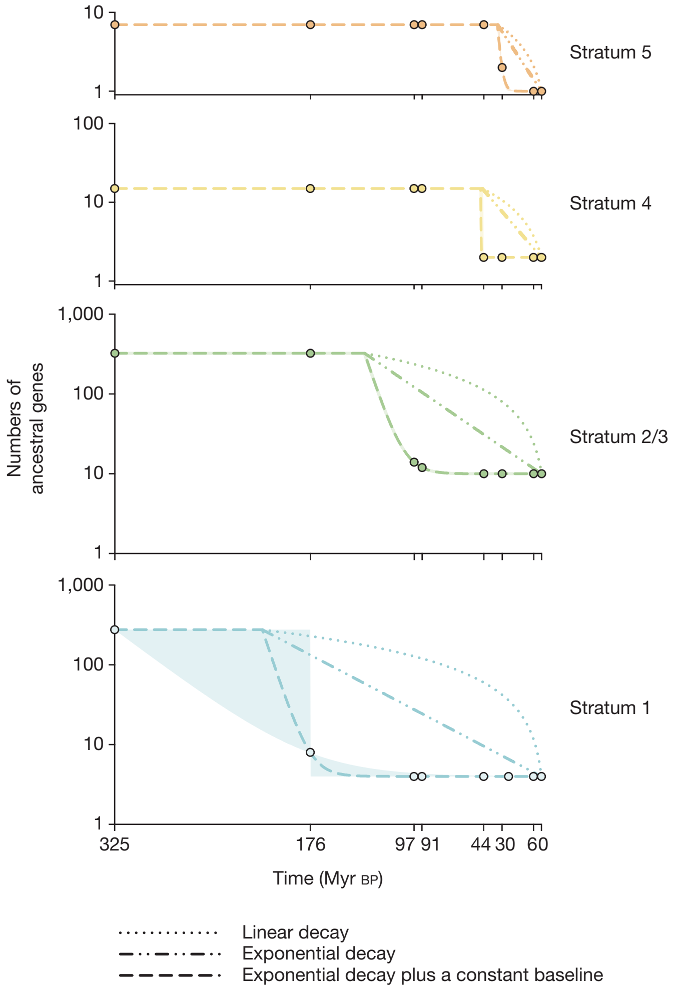


Figure 4 | Decay of Y-linked genes to a baseline level. Gene numbers (on a log scale on the *y* axis) plotted versus time (in Myr before present (Myr BP) on the *x* axis). Filled circles show inferred or observed gene numbers in (from left to right) Ancestral X–Y genes (before stratum formation), the MSY of common ancestor of human and opossum (176 Myr BP), bull (97 Myr BP), mouse and rat (91 Myr BP), marmoset (44 Myr BP), rhesus (30 Myr BP) and chimpanzee (6 Myr BP), and modern human MSY. Lines represent best-fit curves to data points using alternate models of decay. Exponential decay to a constant baseline provides the best fit; shaded regions represent parameters producing an equally good fit.

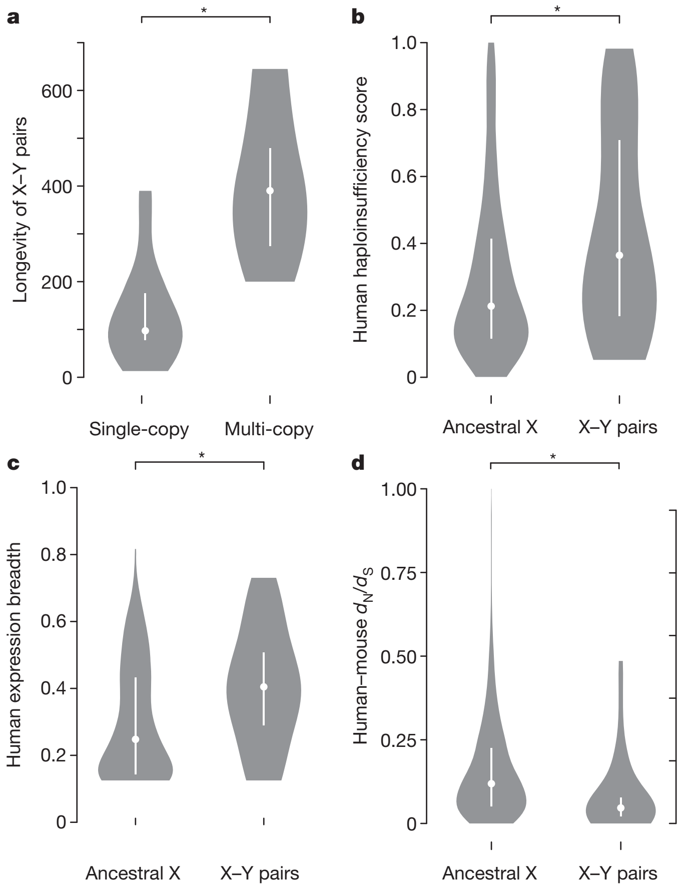


Figure 5 | Factors in the survival of Y-linked genes. Violin plots, white bar, interquartile range; circle, median value; asterisk, significant difference in one-tailed Mann–Whitney *U*-test. **a**, Multi-copy genes ($n = 9$) have greater longevity than single-copy genes ($n = 27$) ($P < 4.28 \times 10^{-5}$). **b**, X–Y pair genes ($n = 32$) have higher haploinsufficiency probability than other ancestral X genes ($n = 478$) ($P < 6.59 \times 10^{-3}$). **c**, X–Y pair genes ($n = 28$) have broader expression across human tissues than other ancestral X genes ($n = 383$) ($P < 2.20 \times 10^{-3}$). **d**, X–Y pair genes ($n = 27$) have lower d_N/d_S ratio than other ancestral X genes ($n = 489$) ($P < 3.39 \times 10^{-4}$).

X–Y pair genes escape X inactivation, but only 168 of 385 remaining ancestral X genes escape (one-tailed Fisher's exact test, $P < 1.89 \times 10^{-3}$) (Supplementary Table 2). The two exceptions, *TSPY* and *RBMY*, are amplified into testis-specific gene families (Extended Data Figs 2 and 6). In mouse, in which X chromosome inactivation is more complete[20], four of nine informative X–Y pair genes escape X inactivation, whereas only five of 344 remaining ancestral genes escape (one-tailed Fisher's exact test, $P < 2.36 \times 10^{-5}$) (Supplementary Table 2). All five exceptions (*Sry, Rbmy, Ube1y, Usp9y* and *Zfy*) evolved testis-specific expression in mouse (Extended Data Fig. 2). Despite differences in the mechanisms of X inactivation between placental and marsupial mammals, all eight informative opossum X–Y pair genes escape X inactivation, but only 15 of 138 remaining ancestral genes escape (one-tailed Fisher's exact test, $P < 1.17 \times 10^{-7}$) (Supplementary Table 2).

The Turner's syndrome phenotype (classically associated with a 45, X karyotype, or monosomy X) suggests a strict dosage requirement for one or more sex-linked genes in humans. If dosage of X–Y pair genes is partially responsible for the Turner's syndrome phenotype, it could explain the differing features of monosomy X in humans and mice. Monosomy X in humans results in poor *in utero* viability. Less than 1 in 100 45,X conceptuses survive to term[25,26]. Those that do survive are often mosaic for all or part of a second sex chromosome[26,27], so that variability in the Turner's syndrome phenotype may reflect variability in dosage of X–Y pair genes among tissues as well as individuals. The mouse phenotype of monosomy X is less severe; animals are small but viable and have reduced fertility[28–30]. This milder phenotype may reflect a dearth of genes on the mouse X chromosome that require two doses: only nine ancestral genes survive on the mouse Y chromosome (compared to 17 in human), and fewer X-linked genes escape inactivation.

Finally, human X-linked intellectual disability syndromes provide evidence for the dosage sensitivity of specific X–Y pair genes. *UTX* (also known as *KMD6A*), *KDM5C* and *NLGN4X* all have Y homologues, escape X inactivation, and appear to be haploinsufficient (Supplementary Table 2). Mutations in *UTX* cause Kabuki syndrome; both duplications and deletions result in multiple congenital anomalies and intellectual disability in males and females[31]. *KDM5C* is associated with X-linked intellectual disability in hemizygous males, and heterozygous females with mild intellectual disability have been reported in several families[32]. In both hemizygous males and heterozygous females, *NLGN4X* mutations are associated with autism spectrum disorders and learning disabilities reminiscent of the cognitive and behavioural phenotypes of Turner's syndrome[33]. Even the human X-homologues of X–Y gene pairs identified only in the opossum (*HCFC1*, *HUWE1* and *MECP2*) still display acute sensitivity to gene dosage. In humans each of these X-linked genes has no Y homologue and is subject to X inactivation[19] (Supplementary

Table 2). Nevertheless, a non-coding mutation causing overexpression of *HCFC1*, as well as duplications of *HUWE1* and *MECP2*, have been implicated in X-linked intellectual disability in human males[34–36]. Thus, even though the human Y-homologues of *HCFC1*, *HUWE1* and *MECP2* were lost and the surviving X-homologues have evolved dosage compensation, their gene dosage remains tightly constrained.

X–Y pair genes functioning across many tissues and cell types may face additional selective constraints that prevent both loss of the Y-linked gene and evolution of a dosage-compensated gene on the X chromosome. In all eight species, single-copy Y-linked genes are broadly expressed across adult tissues (Extended Data Fig. 2), with two major exceptions, in which both members of the X–Y pair share ancestrally restricted expression. *AMELY*, whose orthologue disappeared in the toothless avian lineage, is expressed only in developing tooth buds[37]; and *HSFY*, which is testis specific, and has a chicken orthologue that is predominantly expressed in testis. In chicken, the autosomal orthologues of mammalian X–Y pair genes have significantly broader expression across adult tissues than do the orthologues of ancestral genes that survive only on the X chromosome, and X–Y pair genes maintain this broader expression across mammals (one-tailed Mann–Whitney *U*-test, chicken $P < 3.38 \times 10^{-3}$; human $P < 2.20 \times 10^{-3}$; rhesus $P < 1.39 \times 10^{-7}$; mouse $P < 4.74 \times 10^{-8}$; rat $P < 4.63 \times 10^{-6}$; bull $P < 1.20 \times 10^{-5}$) (Fig. 5c and Supplementary Table 2). This breadth of expression extends to the earliest stages of development. Relative to other X-linked ancestral genes on the X chromosome, X–Y pair genes are enriched for genes upregulated after the onset of zygotic gene activation in a time course of human, mouse and bovine pre-implantation development (one-tailed Fisher's exact test, human $P < 2.13 \times 10^{-2}$; mouse $P < 5.93 \times 10^{-4}$; bull $P < 1.37 \times 10^{-2}$) (Supplementary Table 2). X–Y pair genes are more broadly expressed than other ancestral genes that survive on the X chromosome, across many tissues and developmental time.

Unlike the testis-expressed multi-copy gene families, the broadly expressed, dosage-sensitive single-copy genes of the Y chromosome cannot avoid genetic decay through intrachromosomal gene conversion, and must rely on purifying selection. Our previous survey of human sequence variation among the single-copy genes on the Y chromosome showed that natural selection operated effectively to preserve the amino acid sequences of Y-linked genes in the human lineage[38]. If X–Y gene pairs are haploinsufficient, alleles that alter the function of the X-linked homologues should be detrimental in both males and females. We examined Ensembl human–mouse orthologue alignment data for evidence that the X-linked homologues of X–Y gene pairs were subject to strong purifying selection[23]. Relative to other ancestral genes on the X chromosome, the X-linked homologues of X–Y gene pairs have a reduced ratio of non-synonymous to synonymous substitution rates (dN/dS) (one-tailed Mann–Whitney *U*-test, $P < 3.39 \times 10^{-4}$) (Fig. 5d). We conclude that these broadly expressed, dosage-sensitive X–Y pair genes are under more intense purifying selection than their neighbours on the X chromosome.

Human Y genes ensure male viability

We conclude that the longevity of many Y-linked genes is due to selection to maintain expression, in males, of dosage-sensitive, broadly expressed X–Y gene pairs at levels comparable to their autosomal ancestors. This model predicts that members of surviving single-copy X–Y gene pairs should be functionally interchangeable. Indeed, the human Y-linked genes *RPS4Y1* and *DDX3Y* are functionally interchangeable with their X homologues *in vitro*[39,40], and although the histone demethylase domain of the mouse Y-linked gene *Uty* appears to be inactive, mouse *Utx* and *Uty* are functionally redundant during mouse embryonic development[41–43].

Previous observations suggest that the selective pressures that maintained these Y-linked genes remain strong in the human lineage; about 99% of human 45,X conceptuses are inviable, and those that survive to term are often mosaic for all or part of a second sex chromosome[25–27]. Therefore, we also conclude that the broadly expressed, dosage-sensitive genes of the human Y chromosome—along with their X-homologues, which escape X chromosome inactivation—are collectively haplolethal. We propose that, as a set, these dozen Y-linked genes are essential for the viability of 46,XY fetuses (Methods and Extended Data Fig. 7). Thus we propose a third organismal function of the human Y chromosome: that it carries single-copy genes that ensure male viability. This is distinct from the human Y chromosome's more widely appreciated roles in testis determination through *SRY* and sperm production through ampliconic gene families.

Sex differences in health and disease

All of the myriad differences between human males and females—from anatomy to disease susceptibility—arise from differences in the genes of the X and Y chromosomes that appeared as these chromosomes diverged in gene content from their autosomal ancestors. Of the 17 surviving ancestral genes on the human Y chromosome, four (*SRY*, *RBMY*, *TSPY*, and *HSFY*) have clearly diverged in function from their X homologues (*SOX3*, *RBMX*, *TSPX* and *HSFX*) to play male-specific roles in reproductive development or gametogenesis. Because all genes on the Y chromosome were exposed to selection only in males, even widely expressed ancestral genes may exhibit subtle functional differences from their X-linked homologues. Particularly worthy of consideration are eight global regulators of gene activity that exist as X-encoded and Y-encoded (male-specific) protein isoforms in diverse human tissues: UTX/UTY, EIF1AX/EIF1AY, ZFX/ZFY, RPS4X/RPS4Y1, KDM5C/KDM5D, DDX3X/DDX3Y, USP9X/USP9Y and TBL1X/TBL1Y. These exemplify a fundamental sexual dimorphism, at a biochemical level, throughout the human body, that derives directly from genetic differences between the X and Y chromosomes. It will surely be of interest to determine whether this dimorphism has a role in diseases, outside the reproductive tract, that occur with greater frequency or severity in males or females.

METHODS SUMMARY

We used the SHIMS (single-haplotype iterative mapping and sequencing) strategy to assemble a path of sequenced clones for each organism (Methods). Contigs were ordered and oriented by radiation hybrid mapping using RHMAPPER 1.22 (ref. 44) and extended metaphase and interphase FISH, as previously described[45].

We validated transcription of predicted genes by reverse-transcriptase polymerase chain reaction and capillary sequencing, as well as 454 sequencing of testis complementary DNA (cDNA), as previously described[5].

We relied on Ensembl[23] version 70 to identify 1:1 orthologues between human, chimpanzee, rhesus, marmoset, mouse, rat and bull X chromosomes, the opossum X chromosomes and autosomes, and chicken autosomes, but manually reviewed cases where simple 1:1 orthologues were not clear[2,4]. Within each stratum, we identified X–Y pair genes as ancestral if their X-linked or autosomal orthologues were syntenic in an outgroup.

For each species, we aligned each X–Y pair and calculated d_N, d_S, and d_N/d_S using PAML[46] to identify evolutionary strata. For cross-species phylogenetic analysis, we generated multiple alignments in MUSCLE[47] and used these alignments to generate a tree with 100 bootstrap replicates using DNAML in PHYLIP[48]. Within each stratum, we modelled gene loss as previously described[6].

To calculate longevity, we summed all branch lengths in the most parsimonious tree from each of the species where a gene is present to the last common ancestor before stratum formation.

We mapped published functional annotation data[18–23] onto our set of ancestral genes. We identified pre-implantation expressed genes as previously described[22]. For expression breadth[21], we normalized the expression of each X-linked gene to the highest reads per kilobase per million reads (RPKM) in any tissue, and took the average expression across all tissues. We used PANTHER[49] to calculate the enrichment of Gene Ontology terms in X–Y gene pairs relative to the ancestral X, and used UniProt annotations[50] to identify X–Y pair genes involved in regulatory processes.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 November 2013; accepted 6 March 2014.

1. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286,** 964–967 (1999).

2. Bellott, D. W. *et al.* Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* **466,** 612–616 (2010).
3. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423,** 825–837 (2003).
4. Mueller, J. L. *et al.* Independent specialization of the human and mouse X chromosomes for the male germline. *Nature Genet.* **45,** 1083–1087 (2013).
5. Hughes, J. F. *et al.* Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483,** 82–86 (2012).
6. Hughes, J. F. *et al.* Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463,** 536–539 (2010).
7. Hughes, J. F. *et al.* Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* **437,** 100–103 (2005).
8. Lahn, B. T. & Page, D. C. Functional coherence of the human Y chromosome. *Science* **278,** 675–680 (1997).
9. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434,** 325–337 (2005).
10. Watson, J. M., Spencer, J. A., Riggs, A. D. & Graves, J. A. The X chromosome of monotremes shares a highly conserved region with the eutherian and marsupial X chromosomes despite the absence of X chromosome inactivation. *Proc. Natl Acad. Sci. USA* **87,** 7125–7129 (1990).
11. Murtagh, V. J. *et al.* Evolutionary history of novel genes on the tammar wallaby Y chromosome: implications for sex chromosome evolution. *Genome Res.* **22,** 498–507 (2012).
12. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22,** 2971–2972 (2006).
13. Fisher, R. A. The evolution of dominance. *Biol. Rev. Camb. Philos. Soc.* **6,** 345–368 (1931).
14. Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423,** 873–876 (2003).
15. Kaiser, V. B., Zhou, Q. & Bachtrog, D. Nonrandom gene loss from the *Drosophila miranda* neo-Y chromosome. *Genome Biol. Evol.* **3,** 1329–1337 (2011).
16. Jegalian, K. & Page, D. C. A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature* **394,** 776–780 (1998).
17. Ohno, S. *Sex Chromosomes and Sex-linked Genes* (Springer-Verlag, 1967).
18. Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* **6,** e1001154 (2010).
19. Carrel, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434,** 400–404 (2005).
20. Yang, F., Babak, T., Shendure, J. & Disteche, C. M. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res.* **20,** 614–622 (2010).
21. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338,** 1593–1599 (2012).
22. Xie, D. *et al.* Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res.* **20,** 804–815 (2010).
23. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42,** D749–D755 (2014).
24. Wang, X., Douglas, K. C., Vandeberg, J. L., Clark, A. G. & Samollow, P. B. Chromosome-wide profiling of X-chromosome inactivation and epigenetic states in fetal brain and placenta of the opossum, *Monodelphis domestica*. *Genome Res.* **24,** 70–83 (2014).
25. Cockwell, A., MacKenzie, M., Youings, S. & Jacobs, P. A cytogenetic and molecular study of a series of 45,X fetuses and their parents. *J. Med. Genet.* **28,** 151–155 (1991).
26. Hook, E. B. & Warburton, D. The distribution of chromosomal genotypes associated with Turner's syndrome: livebirth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism. *Hum. Genet.* **64,** 24–27 (1983).
27. Hassold, T., Benham, F. & Leppert, M. Cytogenetic and molecular analysis of sex-chromosome monosomy. *Am. J. Hum. Genet.* **42,** 534–541 (1988).
28. Burgoyne, P. S., Tam, P. P. & Evans, E. P. Retarded development of XO conceptuses during early pregnancy in the mouse. *J. Reprod. Fertil.* **68,** 387–393 (1983).
29. Burgoyne, P. S. & Baker, T. G. Oocyte depletion in XO mice and their XX sibs from 12 to 200 days *post partum*. *J. Reprod. Fertil.* **61,** 207–212 (1981).
30. Burgoyne, P. S., Evans, E. P. & Holland, K. XO monosomy is associated with reduced birthweight and lowered weight gain in the mouse. *J. Reprod. Fertil.* **68,** 381–385 (1983).
31. Lindgren, A. M. *et al.* Haploinsufficiency of KDM6A is associated with severe psychomotor retardation, global growth restriction, seizures and cleft palate. *Hum. Genet.* **132,** 537–552 (2013).
32. Rujirabanjerd, S. *et al.* Identification and characterization of two novel *JARID1C* mutations: suggestion of an emerging genotype-phenotype correlation. *Eur. J. Hum. Genet.* **18,** 330–335 (2010).
33. Lawson-Yuen, A., Saldivar, J. S., Sommer, S. & Picker, J. Familial deletion within NLGN4 associated with autism and Tourette syndrome. *Eur. J. Hum. Genet.* **16,** 614–618 (2008).
34. Huang, L. *et al.* A noncoding, regulatory mutation implicates *HCFC1* in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **91,** 694–702 (2012).
35. Ramocki, M. B., Tavyev, Y. J. & Peters, S. U. The *MECP2* duplication syndrome. *Am. J. Med. Genet. A.* **152A,** 1079–1088 (2010).
36. Froyen, G. *et al.* Copy-number gains of *HUWE1* due to replication- and recombination-based rearrangements. *Am. J. Hum. Genet.* **91,** 252–264 (2012).
37. Lau, E. C., Mohandas, T. K., Shapiro, L. J., Slavkin, H. C. & Snead, M. L. Human and mouse amelogenin gene loci are on the sex chromosomes. *Genomics* **4,** 162–168 (1989).
38. Rozen, S., Marszalek, J. D., Alagappan, R. K., Skaletsky, H. & Page, D. C. Remarkably little variation in proteins encoded by the Y chromosome's single-copy genes, implying effective purifying selection. *Am. J. Hum. Genet.* **85,** 923–928 (2009).
39. Watanabe, M., Zinn, A. R., Page, D. C. & Nishimoto, T. Functional equivalence of human X- and Y-encoded isoforms of ribosomal protein S4 consistent with a role in Turner syndrome. *Nature Genet.* **4,** 268–271 (1993).
40. Sekiguchi, T., Iida, H., Fukumura, J. & Nishimoto, T. Human DDX3Y, the Y-encoded isoform of RNA helicase DDX3, rescues a hamster temperature-sensitive ET24 mutant cell line with a DDX3X mutation. *Exp. Cell Res.* **300,** 213–222 (2004).
41. Welstead, G. G. *et al.* X-linked H3K27me3 demethylase Utx is required for embryonic development in a sex-specific manner. *Proc. Natl Acad. Sci. USA* **109,** 13004–13009 (2012).
42. Shpargel, K. B., Sengoku, T., Yokoyama, S. & Magnuson, T. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *PLoS Genet.* **8,** e1002964 (2012).
43. Lee, S., Lee, J. W. & Lee, S. K. UTX, a histone H3-lysine 27 demethylase, acts as a critical switch to activate the cardiac developmental program. *Dev. Cell* **22,** 25–37 (2012).
44. Slonim, D., Kruglyak, L., Stein, L. & Lander, E. Building human genome maps with radiation hybrids. *J. Comput. Biol.* **4,** 487–504 (1997).
45. Saxena, R. *et al.* The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nature Genet.* **14,** 292–299 (1996).
46. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13,** 555–556 (1997).
47. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
48. Felsenstein, J. PHYLIP - phylogeny inference package (version 3.2). *Cladistics* **5,** 164–166 (1989).
49. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13,** 2129–2141 (2003).
50. The UniProt Consortium Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* **41,** D43–D47 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank W. J. Murphy, E. Owens and J. E. Womak for generating radiation hybrid panels and for assistance in mapping; L. Lyons and W.J.M. for providing the rhesus radiation hybrid panel; A. Kaur for a rhesus cell line; S. Austad, P. Hornsby and S. Tardif for marmoset cell lines; M. Brown for rat cell lines; J.E.W. for bull fibroblasts; W. Johnson and S. O'Neil for rhesus tissues samples; W.J., S.O. and S.T. for marmoset tissue samples; M. Turner for rat tissue samples; J.E.W. for bull tissue samples; P. Samollow for opossum tissue samples; E. Vallender for Tamarin DNA; B. Chowdhary and T. Raudsepp for FISH experiments in the bull; C. Friedman and B. Trask for flow-sorted marmoset Y chromosomes; B.T. for sizing rat Y chromosomes; C. Burge for permission to assemble transcriptome data from SRR594455, SRR594463 and SRR594508; J. Alföldi for permission to assemble transcriptome data from SRR500909; R.B. Norgren for permission to assemble transcriptome data from SRR544870; and A. Godfrey, Y. Hu and B. Lesch for comments on the manuscript. Supported by National Institutes of Health and Howard Hughes Medical Institute.

Author Contributions D.W.B., J.F.H., H.S., S. Rozen, W.C.W., R.A.G., R.K.W. and D.C.P. planned the project. J.F.H., H.S., L.G.B., T.-J.C., N.K. and S.Z. performed BAC mapping, radiation hybrid mapping and real-time polymerase chain reaction analyses. T.G., S. Rock, C.K., R.S.F., S.D., Y.D., D.M., Z.K., L.L., C.B., Q.W., J.W., M.H., S.L., L.N. and D.M.M. were responsible for BAC sequencing. D.W.B., J.F.H. and H.S. performed comparative sequence analyses. T.P. performed FISH analyses. D.W.B. and D.C.P. wrote the paper.

Author Information The cDNA sequences of Y-linked genes and their X-linked homologs have been deposited in GenBank (http://www.ncbi.nlm.nih.gov) under accession numbers FJ526999–FJ527008, FJ627275, FJ627276, FJ627278, FJ659845, FJ959389, GQ253467–GQ253475, GQ338825, GU304599–GU304603, GU304606, GU304607, JF487792–JF487795, JF827151, JF827152, JN086997, JN585955, JN585956, JQ313990–JQ313992 and BioProject PRJNA221163. The 454 and Illumina testis cDNA sequences have been deposited in GenBank under accession numbers SRX335333, SRX335335, SRX335470, SRX335472, SRX335475–SRX335477, SRX358238 and SRX359414. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.W.B. (bellott@wi.mit.edu).

METHODS

Single-haplotype iterative mapping and sequencing (SHIMS) strategy. The single-haplotype iterative mapping and sequencing (SHIMS) strategy was used to assemble partial male-specific region of the Y (MSY) sequences for marmoset, mouse, rat, bull and opossum. We previously employed the SHIMS strategy to obtain the full-length MSY sequences of human, chimpanzee and rhesus macaque[5,6,51]. The major steps in the SHIMS strategy are outlined below:

Initial BAC selection and sequencing. MSY-derived bacterial artificial chromosome (BAC) clones are identified and organized into contigs of overlapping BACs using one or more of the following methods based on resource availability: (1) high-density filter hybridization using pools of overgo probes, (2) electronic mapping of BAC-end sequences to female genomic sequence and (3) BAC fingerprint contig analysis. Assembled MSY contigs are verified by PCR using MSY-specific STS markers. Tiling paths of clones are selected for sequencing.

Distinguishing repeat copies and finding true tiling paths. Overlaps between BACs within repetitive regions are scrutinized for sequence differences or sequence family variants (SFVs). If SFVs are found, this indicates that the BACs belong to distinct copies of the same repeat unit. SFV patterns are then used to identify true overlapping BACs. New tiling paths are produced, and the process is reiterated until all overlaps are consistent.

Extension and joining of BAC contigs. Identify clones that extend outward from or link existing contigs using high-density filter hybridization.

Clone selection. We designed overgo probes from male-specific sequences identified by electronic subtraction of female genomic sequences from male (or mixed male and female) genomic sequences. Because of this approach, our clone selection was not biased towards gene-containing regions. We selected clones from existing male BAC libraries CHORI-259, RPCI-24, CHORI-240, and VMRC-6 (http://bacpac.chori.org), as well as custom BAC libraries MARMAEX, RNAEX, RNECO, BTDAEX and MDAEX constructed by Amplicon Express (http://www.genomex.com).

Finished sequence quality. For each large-insert BAC or fosmid clone, the consensus sequence was completed to a minimum standard of double strand coverage of each base, with a minimum of 2 clones and 2 reads with no ambiguities or 2 sequencing chemistries with a minimum of 2 clones and 3 reads with no ambiguities, or covered by high quality (phred quality ≥ 30). The assembly consensus was derived from Sanger or 454 sequencing platforms supplemented with Illumina or SOLiD data. We attempted to resolve all sequencing problems, such as compressions and repeats. All regions were covered by sequence from more than one subclone. We attempted to close all gaps, and where gaps remain, all contigs were ordered and oriented. Each assembly was confirmed by restriction digest. All bases where the sequence quality does not meet the standard for finished sequence are indicated in the annotation of the GenBank record.

Sequencing error rate. The sequencing error rate for the partial MSY sequences for marmoset, mouse, rat, bull and opossum is approximately one nucleotide per 0.3 Mb.

Order and orientation of contigs. We ordered and oriented our clone-based contigs using both radiation hybrid mapping using RHMAPPER 1.22 (ref. 44) and extended metaphase and interphase fluorescence *in situ* hybridization (FISH), as previously described[45]. We used a previously published 10,000-rad rhesus macaque radiation hybrid panel[52], and a set of new 25,000-rad radiation hybrid panels from marmoset, mouse, bull and opossum, constructed by William J. Murphy, James E. Womack and Elaine Owens. For bull FISH, we used a primary fibroblast cell line derived from the sequenced animal, L1 Domino (JEW 85), received from James E. Womack and Elaine Owens of Texas A&M University. For marmoset FISH, we used cell lines WHT5952 (father of sequenced animal) and WHT5955 (brother of sequenced animal) received from Suzette Tardif and Peter Hornsby in the Sam and Ann Barshop Institute for Longevity and Ageing Studies at the University of Texas Health Science Center. For rat FISH, we created cell line WHT5890, embryonic fibroblasts derived from non-phenotypic SHR rat line from Charles River Labs. For mouse FISH, we established embryonic fibroblast cell lines from the C57BL/6 strain from Jackson Laboratories. For opossum FISH, we used primary fibroblast cell line WHT6354 derived from opossum A0067 from Paul Samollow of Texas A&M University.

Gap closure. Regions composed of repeats with units less than 10 kb and greater than 99% identity frustrate the assembly of individual BAC clones and are not well-represented in our assemblies. These regions include both gene-poor regions like centromeres, telomeres and heterochromatin, as well as gene-rich regions, such as the *TSPY* arrays on the human and bull Y chomosomes. No current technology is able to access these regions. Wherever possible we attempted to find the boundaries of these arrays, obtain a representative repeat unit, and verify the contiguity of the array by FISH.

The gaps in both bull and opossum assemblies (Extended Data Fig. 1) are the result of arrays of short, highly identical repeats of this type. The bull Y-chromosome assembly is interrupted by extremely long tandem arrays of a ~3 kb repeat unit, but all contigs are ordered and oriented, and the homogeneity of these arrays was confirmed by FISH. The opossum Y-chromosome assembly is interrupted by stretches of several different heterochromatic repeat units. The opossum Y chromosome is too small to resolve these regions by FISH. However, we are confident that our assembly is not biased towards gene-rich regions due to our almost exclusive use of electronic subtraction to generate probes.

Annotation of ancestral genes. We validated transcription of predicted genes by reverse-transcriptase polymerase chain reaction and capillary sequencing, as well as 454 sequencing of testis complementary DNA (cDNA), as previously described[5].

We relied on Ensembl[23] version 70 to identify 1:1 orthologues between human, chimpanzee, rhesus, marmoset, mouse, rat and bull X chromosomes, the opossum X chromosomes and autosomes, and chicken autosomes, but manually reviewed cases where simple 1:1 orthologues were not clear[2,4]. Within each stratum, we identified X–Y pair genes as ancestral if their X-linked or autosomal orthologues were syntenic in an outgroup.

We mapped published functional annotation data[18–23] onto our set of ancestral genes. We identified pre-implantation expressed genes as previously described[22]. For expression breadth[21], we normalized the expression of each X-linked gene to the highest reads per kilobase per million reads (RPKM) in any tissue, and took the average expression across all tissues. We used UniProt annotations[50] to identify X–Y pair genes involved in regulatory processes.

Identification and modelling of evolutionary strata. For each species, we aligned each X–Y pair and calculated d_N, d_S, and d_N/d_S using PAML[46] to identify evolutionary strata. For cross-species phylogenetic analysis, we generated multiple alignments in MUSCLE[47] and used these alignments to generate a tree with 100 bootstrap replicates using DNAML in PHYLIP[48]. Within each stratum, we modelled gene loss as previously described[6].

To calculate longevity, we summed all branch lengths in the most parsimonious tree from each of the species where a gene is present to the last common ancestor before stratum formation.

PANTHER statistical overrepresentation test. We employed the PANTHER statistical overrepresentation test[49] to identify functional coherence among the 36 ancestral X–Y pair genes relative to the remaining ancestral X genes. For each functional category, the PANTHER software employs a binomial test to identify statistically significant over-representation (or under-representation) of the genes in an input list relative to the genes in a reference list[53]. This test makes no assumptions about the processes that generated either the input or reference gene lists, aside from the null hypothesis that both the input and reference list are drawn from the same population, such that each functional category is equally well represented in the two lists[53].

We manually curated our gene lists to ensure that any over-representation we identified was the result of processes that favoured the survival of ancestral genes on the Y chromosome, rather than the processes that drove gene acquisition and amplification. First, we restricted our analyses to X–Y gene pairs that included one of the 639 ancestral X-linked genes we identified in our reconstruction of the ancestral autosomes from which the X and Y chromosomes evolved (Supplementary Table 2). Second, we excluded any X–Y gene pairs we could identify as arising from gene acquisition by the Y chromosome after the start of decay; for example, we excluded the X–Y pair genes resulting from the human-specific X-transposed region.

Out of the 639 ancestral X-linked genes, we identified 36 with Y homologues (Fig. 1) that appear to have survived through the genetic decay of the Y chromosome in any one of our 8 species. All 36 of these genes mapped to a human identifier in PANTHER. Of the 603 remaining ancestral genes, 11 were lost in the human lineage, and 38 did not map to a human identifier in PANTHER, leaving 554 ancestral X genes without a surviving Y homologue in any of our 8 species (Supplementary Table 2).

We used the PANTHER statistical overrepresentation test to identify functional annotations that were enriched among the 36 ancestral X–Y pair genes that survive on the Y chromosome of one or more of the eight species we sequenced, relative to the reference list of 554 other ancestral X genes (Extended Data Table 1). We selected the 554 other ancestral X genes as a reference list, instead of all human genes, to control for any functional coherence among the ancestral genes that pre-dated the start of Y-chromosome decay, as well as the possibility that the annotation of the X chromosome is more complete than that of the autosomes.

We found that the annotation of the combined set of 590 ancestral X genes (36 ancestral X–Y pairs and 554 other ancestral X genes) is more complete than the rest of the human genome. Relative to all human genes, the 590 ancestral X genes are significantly underrepresented for genes that are "Unclassified" in the GO Biological Process ($P < 1.96 \times 10^{-7}$), GO Molecular Function ($P < 1.52 \times 10^{-2}$), and Panther Protein Class ($P < 1.00 \times 10^{-6}$) categories (Supplementary Table 4). On the other hand, the 590 ancestral X genes are over-represented for three GO Biological Process annotations: "neurological system process" ($P < 3.14 \times 10^{-2}$), "cellular process" ($P < 4.50 \times 10^{-2}$), and "synaptic transmission" ($P < 4.59 \times 10^{-2}$) (Supplementary Table 4). We note that the "cellular process" annotation encompasses "synaptic transmission," and that "cellular process" would not reach statistical significance if genes annotated as "synaptic transmission" were excluded. We obtained similar results when we excluded the 36 X–Y gene pairs and tested the 554 other ancestral X genes against all human genes, although the "Unclassified" annotation in the GO Molecular Function category failed to reach significance (Supplementary Table 4). We interpret these results as evidence that the intensive study of X-linked intellectual disability syndromes has produced a richer annotation of brain and cognitive functions on the X chromosome relative to the autosomes.

Identification and recalibration of evolutionary strata. We identified chromosomal fusions and evolutionary strata across our tree of species, using a combination of information: syntenic orthologues across species, synonymous nucleotide divergence between X–Y gene pairs, and phylogentic analysis of X–Y gene pairs.

A chromosomal fusion in the ancestor of placental mammals. Previous comparisons between marsupial and placental sex chromosomes identified a conserved region shared between the sex chromosomes of placental and marsupial mammals, and an added region unique to the sex chromosomes of placental mammals[10]. Orthologues of genes from the added and conserved regions are found on separate autosomes in the chicken genome, the best assembled outgroup to placental and marsupial mammals, as well as in the genomes of 4 teleost fish[2,9]. These inter-species comparisons of X chromosomal and autosomal gene content established the model that the present day human X and Y chromosomes derived from the X-conserved region existed in the common ancestor of placental and marsupial mammals, and later, a chromosomal fusion brought the added and conserved regions together in the ancestor of placental mammals.

Our comparisons of Y-linked gene content support this model. Across all seven placental mammals, we identified 17 X–Y pairs that derive from the added region (Fig. 1). As, expected, none of these pairs have an orthologue on the opossum Y chromosome (Fig. 1). Additionally, we note that the opossum orthologues of placental added region genes reside on two autosomes in opossum, chromosome 4 and chromosome 7 (Supplementary Table 2 and Extended Data Fig. 3). Because the orthologues of placental X-added region genes are also syntenic in an outgroup, chicken[2,9], we conclude that the ancestral autosome orthologous to the added region of the placental sex chromosomes broke apart in the opossum lineage (Fig. 3).

Reconstruction of evolutionary strata. The chromosomal fusion event recorded in the placental added and conserved regions served as a palimpsest for the formation of evolutionary strata. Previous comparisons of the human X and Y chromosomes identified five evolutionary strata overlaid across the added and conserved regions on the X chromosome[1,9]. The oldest evolutionary strata, stratum one and stratum two, occupied the X-conserved region, whereas the X-added region contained strata three, four, and five, as well as the freely recombining pseudoautosomal region (PAR)[1,9]. We re-examined these findings across our expanded set of species and gene pairs. Within each species, we aligned single-copy X–Y gene pairs and calculated the nucleotide divergence (d_S) between them (Supplementary Table 5). In the two oldest strata, uncertainty in the levels of divergence prevented us from distinguishing strata, in these cases we sought to distinguish strata by phylogenetic analysis (Extended Fig. 4). The data from our broader comparison provides additional details that allow us to refine previous reconstructions of the evolutionary trajectory of the human sex chromosomes. In particular, we find no support of the distinction between strata two and three, and propose that a single combined stratum arose in the placental lineage after the fusion of the added and conserved regions.

Stratum two formed independently in placental and marsupial lineages. Based on the analysis of five X–Y gene pairs, previous reconstructions placed the two oldest strata before the divergence of placental and marsupial mammals[1,3]. We found that placental Y-linked genes from both stratum one and stratum two have orthologues in the opossum (Fig. 1), as would be expected if both strata formed in the common ancestor of placental and marsupial mammals. Alternatively, the survival of Y-linked genes in both lineages could be the result of independent stratum formation and convergent survival of Y-linked genes after the divergence of marsupial and placental mammals. We examined both possibilities in light of our new data from the marsupial lineage. Sixteen opossum X–Y pairs are drawn from across the entire X-conserved region, encompassing both stratum one and stratum two. However, all opossum X–Y pairs (with the exception of *SOX3*/*SRY*) displayed a similarly high level of divergence ($d_S >= 1$) (Supplementary Table 5).

Because saturation for synonymous substitutions prevented us from using nucleotide divergence to distinguish these ancient strata in the opossum, we sought to distinguish between them by phylogenetic analysis of X–Y gene pairs across all eight species, using autosomal orthologues in chicken as the outgroup. We found that across both placental and marsupial mammals, orthologues of the stratum one genes *SRY*, *RBMY* and *HSFY* were more closely related to each other than to X-linked homologues (Extended Data Fig. 4). Genes from stratum two showed a different pattern; as a group, placental orthologues of *UBE1Y* and *KDM5D* are more closely related to placental X-linked homologues than to their marsupial orthologues (Extended Data Fig. 4). We conclude that statum one, containing *SRY*, the male sex-determining gene[54,55], evolved only once, before the divergence of marsupial and placental mammals, but that the formation of a second stratum proceeded independently in both lineages (Figs 1 and 3).

No support for the distinction between stratum two and stratum three. Previous reconstructions drew a distinction between stratum two and stratum three because stratum two had been dated before the divergence of placental and marsupial mammals and stratum three contained genes from the region added to the placental sex chromosomes. After finding that only the first and not the second stratum preceded the divergence of placental and marsupial lineages, we re-examined the distinction between stratum two and stratum three in placental mammals. We compared stratum two and stratum three gene pairs only from the four primate species; no single-copy gene pairs from stratum two survived on the bull Y chromosome, and single-copy gene pairs from both strata are saturated for synonymous substitutions in the rodent lineage (Fig. 1 and Supplementary Table 5). We also excluded *AMELY* and *ZFY*, which participated in interchromosomal gene conversion after stratum formation (Supplementary Table 5 and Extended Data Fig. 5)[56,57]. We found that within each of the four primate species, the divergence between *KDM5C* and *KDM5D* in stratum two is within the range of divergence of X–Y gene pairs from stratum three (Supplementary Table 5). Without phylogenetic or divergence data that distinguish stratum two from stratum three, we propose that together they represent a single stratum (Figs 1 and 3). This combined stratum formed in the ancestor of all placental mammals, after the chromosomal fusion event expanded the PAR of the X and Y chromosomes, but before bull diverged from the other six species, more than 97 millon years ago (Fig. 3)[12].

Location of the ancestral placental PAR boundary. The formation of this combined stratum defined the PAR boundary in the placental ancestor, but subsequent X–Y gene conversion events in *AMELY* have made it difficult to establish the location of this boundary using only data from the human X and Y chromosomes, with proposed boundaries ranging in location from as distal as between *KAL1* and *TBL1X* and as proximal as between *AMELX* and *TMSB4X*[1,3,9,58]. The 4.2 megabases between *KAL1* and *TMSB4X* comprise almost 3% of the human X chromosome. Our expanded data set provides additional constraints that narrow this region by a factor of 10. We find that *AMELY* is present on the human, chimpanzee, rhesus macaque and bull Y chromosomes, while *TBL1Y* is present only in human, rhesus macaque and, as a pseudogene, in chimpanzee (Fig. 1). The bovine orthologue of *TBL1X* is located in the PAR, and furthermore, *MID1*, which is located between *TBL1X* and *AMELX* on the human X chromosome, has an orthologue in the mouse PAR (Extended Data Table 2)[59]. We conclude that the ancestral placental PAR boundary was proximal to both *TBL1X* and *MID1*, but distal to *AMELX*. This places *TBL1Y* in stratum four, and *AMELY* in the combined stratum two/three. The low divergence between *AMELX* and *AMELY* is probably the result of lineage-specific X–Y gene conversion events after stratum formation, similar to what has been observed for *ZFY* (Supplementary Table 5 and Extended Data Fig. 5)[56,57].

Lineage-specific evolutionary strata in primates. After the formation of the stratum that established the ancestral placental PAR boundary, lineage-specific evolutionary strata continued to form. Previous reconstructions identified two additional strata in the human lineage with a boundary between *PRKX* and *NLGN4X*[9]. We recalculated the age of human strata 4 and 5 following previously published methods[9], using the updated figure of 29.6 million years ago for the divergence between old world monkeys and hominoids[12].

NLGN4Y, from stratum four, is present in all four primate species, whereas *TBL1Y* is present in human and rhesus macaque, with a pseudogene in chimpanzee. The X–Y divergence in human stratum four is compatible with an origin in the simian ancestor, over 44 million years ago, close to the time of divergence of platyrhine and catarrhine primates (Fig. 3)[9,12].

In contrast, human stratum five dates to 32–34 million years ago, before the divergence of rhesus macaque from human and chimpanzee[9,12]. All three species share the *PRKY* gene, as well as a common PAR boundary[5]. We conclude that stratum five was already established in the catarrhine ancestor, and afterwards, no further strata formed in the human, chimpanzee and rhesus lineages (Fig. 3), although subsequent insertions, deletions, and rearrangements generated different configurations of the male-specific region of the Y chromosome in each species[5].

Independently, the marmoset lineage also formed a fifth stratum with a more distal PAR boundary than the human, chimpanzee, and rhesus (Fig. 1 and Supplementary Fig. 7). Because the marmoset whole genome shotgun sequence is a mixture of male and female sequence, and this marmoset-specific stratum formed relatively recently, it is not possible to differentiate between X and Y derived contigs in the marmoset whole genome shotgun sequence. *P2RY8Y*, *SFRS17AY*, and *ZBED1Y* are the only survivors out of 24 ancestral genes in this stratum (Fig. 1 and Supplementary Table 2), demonstrating that, at least while strata are young, genetic decay is both swift and extensive[5,60].

Modelling kinetics of Y-chromosome decay. We modelled the numbers of ancestral genes within individual MSY strata as a function of time in millions of years before the present by fitting a one-phase exponential decay model with a baseline constant (below) to our data using nonlinear regression analysis in GraphPad Prism 5.0. Parameters for each stratum are given in the Source Data for Fig. 4. This one-phase exponential decay model gives the number of genes at time t, $N(t)$:

$$N(t) = (N_0 - b)e^{-Kt} + b$$

Where N_0 is the number of genes within a given stratum in ancestral autosomal/pseudoautosomal portion of genome at the start of stratum formation, K is the decay constant, and b is the baseline (approximated by the number of active ancestral genes within that stratum on the human Y chromosome).

Haplolethality of broadly expressed, dosage-sensitive X–Y pair genes. We propose that the broadly expressed, dosage-sensitive genes of the human Y chromosome, along with their X homologues that escape X chromosome inactivation, are collectively haplolethal. Twelve human XY gene pairs meet this criterion: *RPS4X/RPS4Y1, ZFX/ZFY, TBL1X/TBL1Y, PRKX/PRKY, USP9X/USP9Y, DDX3X/ DDX3Y, UTX/ UTY, TMSB4X/ TMSB4Y, NLGN4X/ NLGN4Y, TXLNG/CYORF15, KDM5C/KDM5D* and *EIF1AX/EIF1AY*.

We compiled a list of cases with non-mosaic partial-Y deletions removing one or more of these genes to determine if any single gene was haplolethal. We found that the Y-homologue of each X–Y gene pair was deleted in one or more cases (Extended Data Fig. 7 and Extended Data Table 3). Thus, we attribute the inviability of 45,X conceptuses to a collective haplolethality for several X–Y gene pairs, and not to any single gene pair. Supporting the notion that these gene pairs are dosage-sensitive, *TBL1Y* and *PRKY*, two genes deleted in the rare J2e1*/M241 Y chromosome haplotype[61], are the only 2 of these 12 gene pairs with X-linked homologues that do not always escape X-inactivation[19].

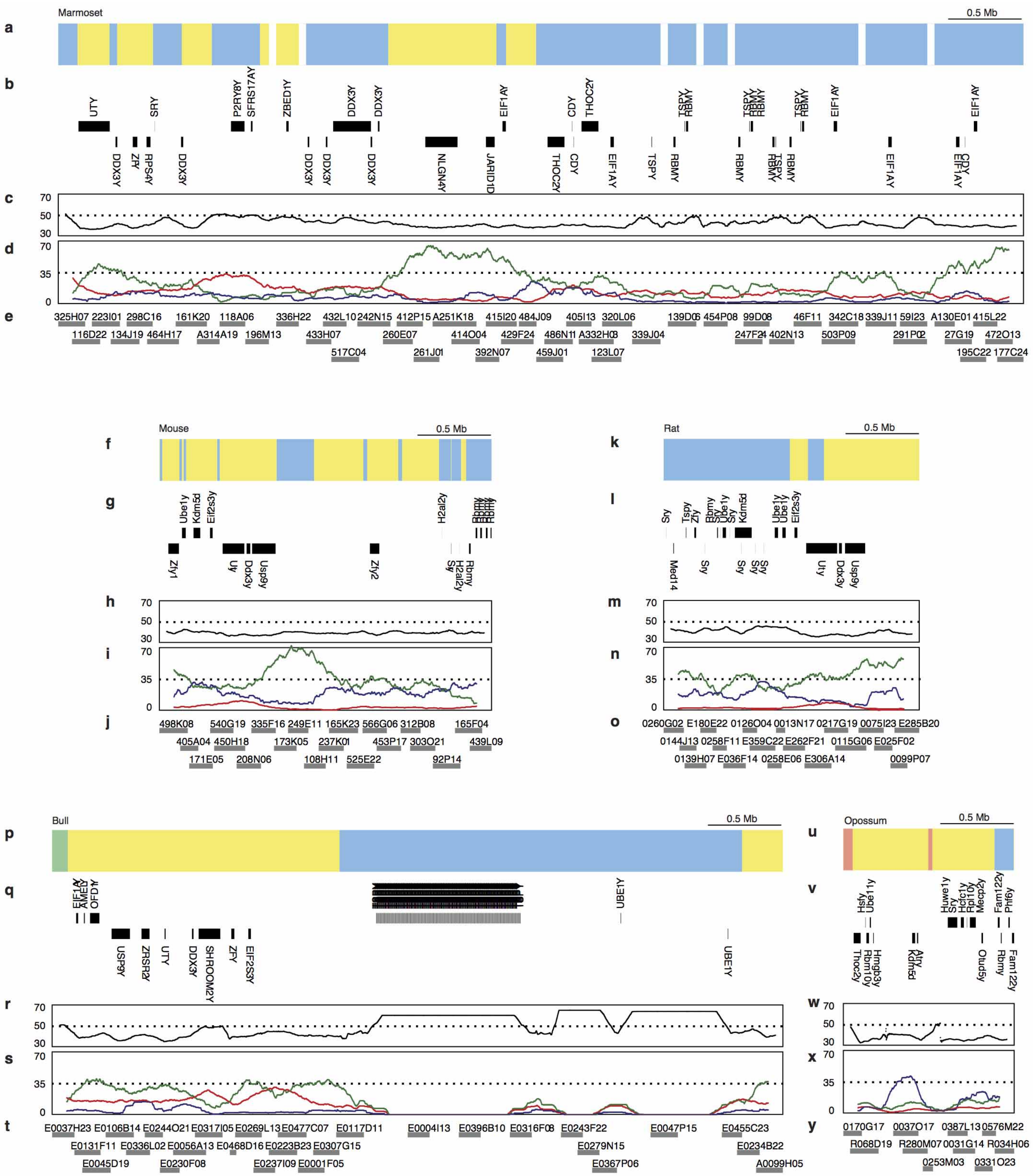
We also searched the literature for reports of structurally variant X chromosomes in females, where one X chromosome was deleted for one or more of these 12 genes (Extended Data Fig. 7 and Extended Data Table 3). These reports are not inconsistent with a collective haplolethality for X–Y gene pairs, but the interpretation of these cases is complicated by viability effects mediated by the X-inactivation centre (XIC), and a possible critical region for ovarian failure near *USP9X*[62].

We found cases where a variant X chromosome has been transmitted from mother to daughter, and which are therefore unlikely to be mosaic, that delete as many as 7 genes (*PRKX*, *NLGN4X*, *TBL1X*, *TMSB4X*, *TXLNG*, *EIF1AX* and *ZFX*)[63–69].

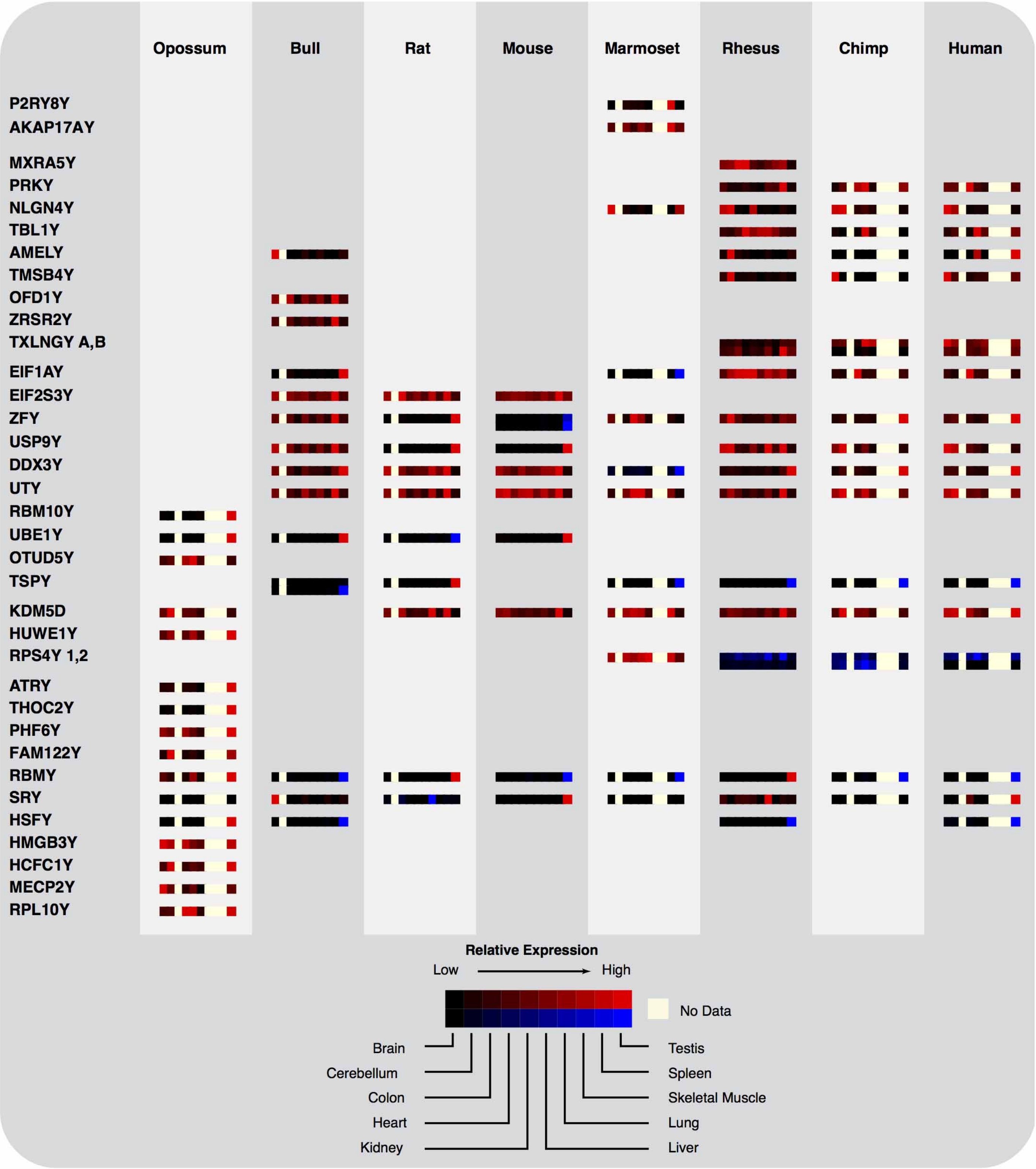
We also found reports of extensive *de novo* deletions that eliminate 11 of these 12 genes, leaving only *RPS4X* on the long arm[66,69]. However, we cannot exclude the possibility that these cases are mosaic for 46,XX cells in a cell lineage other than the blood. The absence of familial cases of deletions of this type may because of a critical region for ovarian failure on the short arm of the X chromosome; both *ZFX* and *USP9X* have been proposed as candidate genes[62].

We could not find any reports of deletions of *RPS4X*. *RPS4X* is located on the long arm, between the centromere and the XIC. We believe that the absence of reports of X chromosome variants deleted for *RPS4X* reflects the proximity of *RPS4X* to the XIC rather than haplolethality of *RPS4X*.

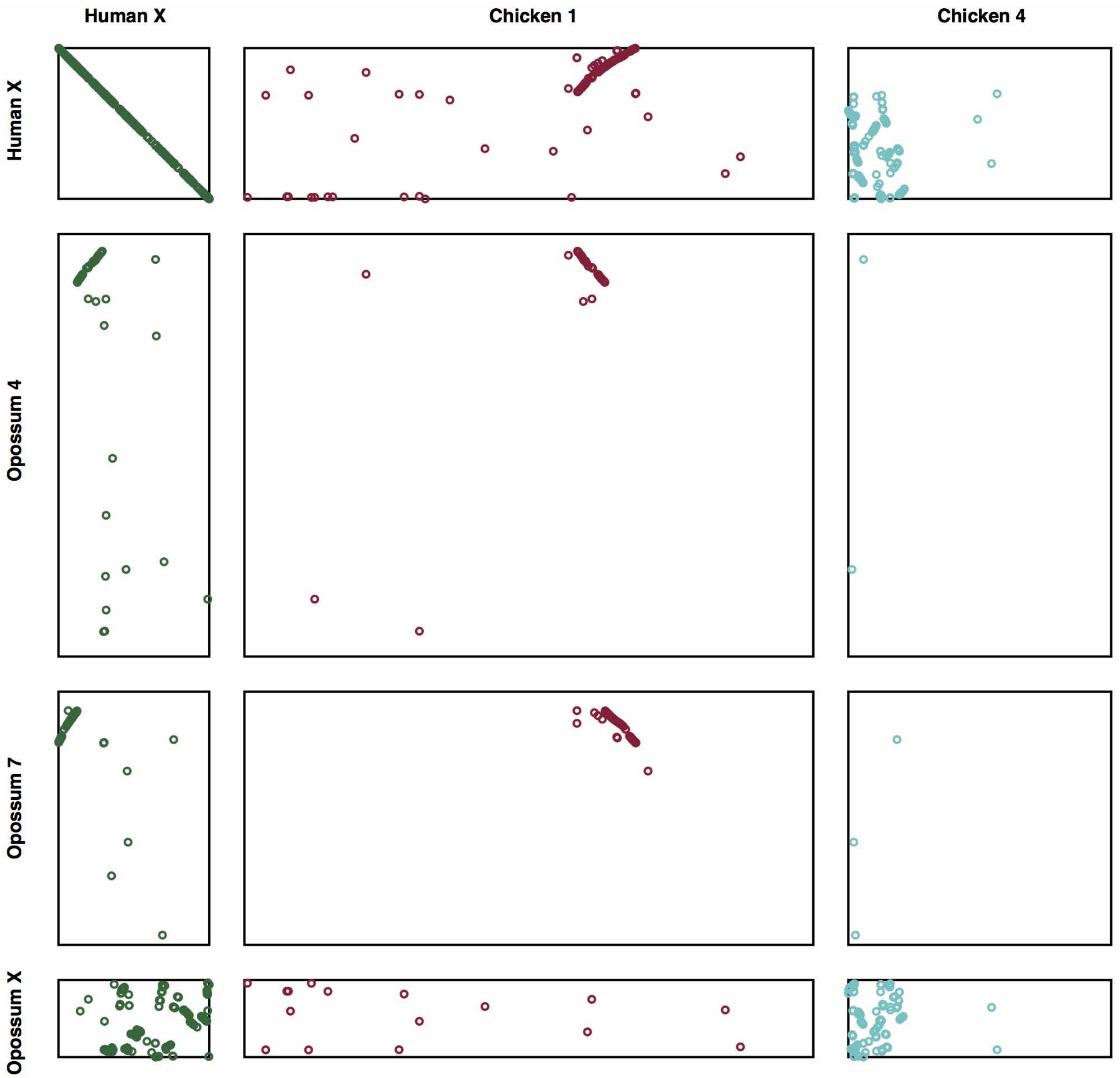
51. Kuroda-Kawaguchi, T. *et al.* The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genet.* **29,** 279–286 (2001).
52. Karere, G. M., Froenicke, L., Millon, L., Womack, J. E. & Lyons, L. A. A high-resolution radiation hybrid map of rhesus macaque chromosome 5 identifies rearrangements in the genome assembly. *Genomics* **92,** 210–218 (2008).
53. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols* **8,** 1551–1566 (2013).
54. Gubbay, J. *et al.* A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346,** 245–250 (1990).
55. Sinclair, A. H. *et al.* A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346,** 240–244 (1990).
56. Hayashida, H., Kuma, K. & Miyata, T. Interchromosomal gene conversion as a possible mechanism for explaining divergence patterns of ZFY-related genes. *J. Mol. Evol.* **35,** 181–183 (1992).
57. Marais, G. & Galtier, N. Sex chromosomes: how X-Y recombination stops. *Curr. Biol.* **13,** R641–R643 (2003).
58. Iwase, M. *et al.* The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species. *Proc. Natl Acad. Sci. USA* **100,** 5258–5263 (2003).
59. Dal Zotto, L. *et al.* The mouse *Mid1* gene: implications for the pathogenesis of Opitz syndrome and the evolution of the mammalian pseudoautosomal region. *Hum. Mol. Genet.* **7,** 489–499 (1998).
60. Bachtrog, D. The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* **179,** 1513–1525 (2008).
61. Jobling, M. A. *et al.* Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing *Amelogenin Y*. *Hum. Mol. Genet.* **16,** 307–316 (2007).
62. Jones, M. H. *et al.* The *Drosophila* developmental gene fat facets has a human homologue in Xp11.4 which escapes X-inactivation and has related sequences on Yq11.2. *Hum. Mol. Genet.* **5,** 1695–1701 (1996).
63. Adachi, M., Tachibana, K., Asakura, Y., Muroya, K. & Del Ogata, T. Del(X)(p21.1) in a mother and two daughters: genotype-phenotype correlation of Turner features. *Hum. Genet.* **106,** 306–310 (2000).
64. Chocholska, S., Rossier, E., Barbi, G. & Kehrer-Sawatzki, H. Molecular cytogenetic analysis of a familial interstitial deletion Xp22.2-22.3 with a highly variable phenotype in female carriers. *Am. J. Med. Genet. A.* **140A,** 604–610 (2006).
65. Good, C. D. *et al.* Dosage-sensitive X-linked locus influences the development of amygdala and orbitofrontal cortex, and fear recognition in humans. *Brain* **126,** 2431–2446 (2003).
66. James, R. S. *et al.* A study of females with deletions of the short arm of the X chromosome. *Hum. Genet.* **102,** 507–516 (1998).
67. Massa, G., Vanderschueren-Lodeweyckx, M. & Fryns, J. P. Deletion of the short arm of the X chromosome: a hereditary form of Turner syndrome. *Eur. J. Pediatr.* **151,** 893–894 (1992).
68. Zinn, A. R. *et al.* Del (X)(p21.2) in a mother and two daughters with variable ovarian function. *Clin. Genet.* **52,** 235–239 (1997).
69. Zinn, A. R. *et al.* Evidence for a Turner syndrome locus or loci at Xp11.2-p22.1. *Am. J. Hum. Genet.* **63,** 1757–1766 (1998).



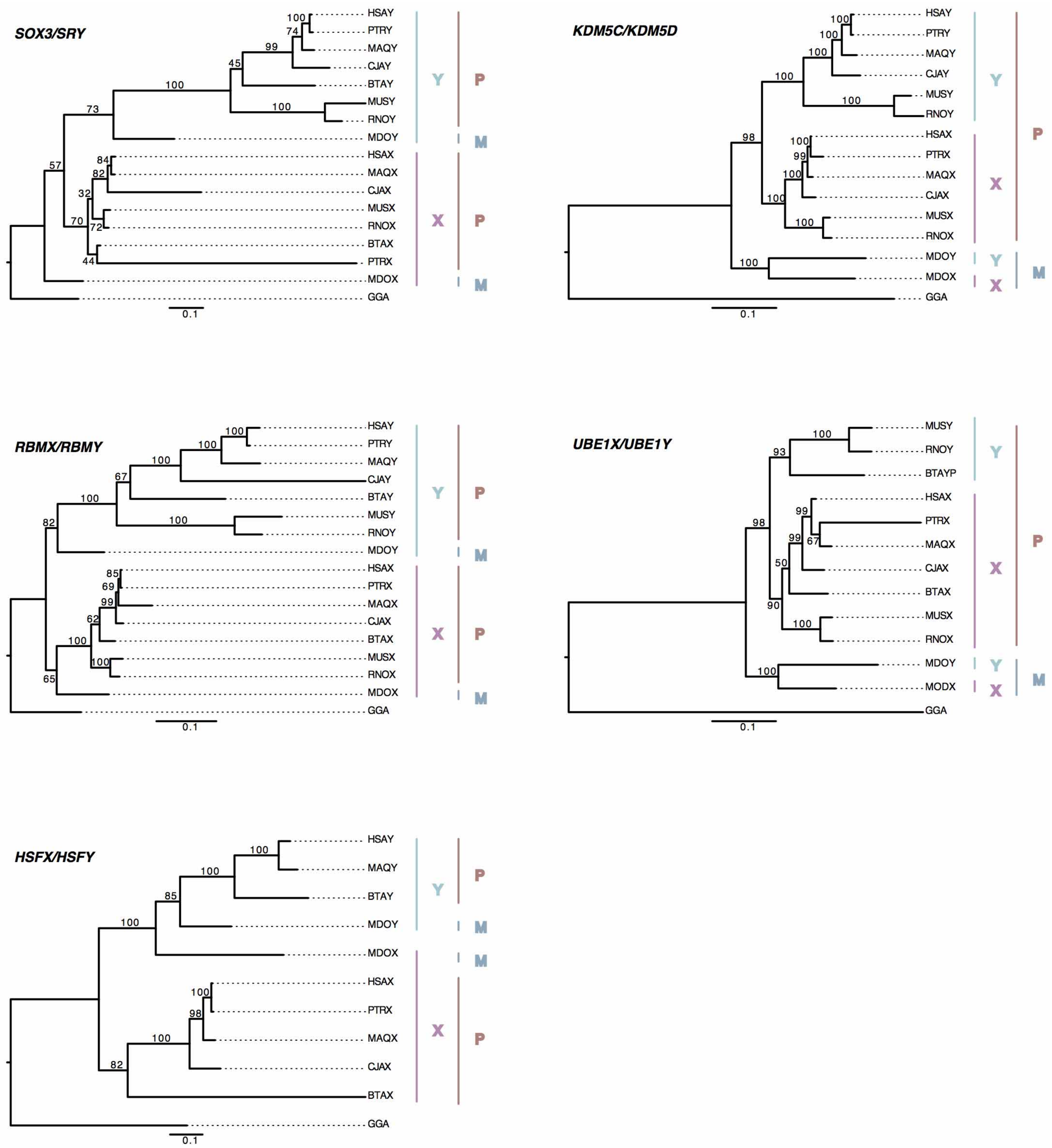
Extended Data Figure 1 | Annotated sequence contigs from the MSY of five species. All sequence features and BACs drawn to scale. **a**–**e**, Marmoset MSY. **f**–**j**, Mouse MSY. **k**–**o**, Rat MSY. **p**–**t**, Bull MSY. **u**–**y**, Opossum MSY. **a**, **f**, **k**, **p**, **u**, Schematic representation of assembled contigs and sequence classes: X-degenerate (yellow); ampliconic (blue); pseudoautosomal (green); heterochromatic (pink). Gaps shown in white. **b**, **g**, **l**, **q**, **v**, Positions of all intact, actively transcribed genes. Plus (+) strand above, minus (−) strand below. **c**, **h**, **m**, **r**, **w**, G + C content (%) calculated in a 100-kb sliding window with 1-kb steps. **d**, **i**, **n**, **s**, **x**, Alu (red), LINE (green), and endogenous retrovirus (blue) content (%) calculated in a 200-kb sliding window with 1-kb steps. **e**, **j**, **o**, **t**, **y**, Sequenced MSY BACs. Each bar represents the size and position of one BAC clone, labelled with the library identifier. **e**, BAC clones with no prefix are from the CHORI-259 library; BAC clones with "A" prefix are from the MARMAEX (Amplicon Express) library. **j**, All BAC clones are from the RPCI-24 library. **o**, BAC clones without prefix are from the RNAEX library; BAC clones with "E" prefix are from the RNECO library (both from Amplicon Express). **t**, BAC clones without prefix are from the CHORI-240 library; BAC clones with "E" prefix are from the BTDAEX library (both from Amplicon Express). **y**, BAC clones with no prefix are from the VMRC6 library; BAC clones with "A" prefix are from the MDAEX (Amplicon Express) library.



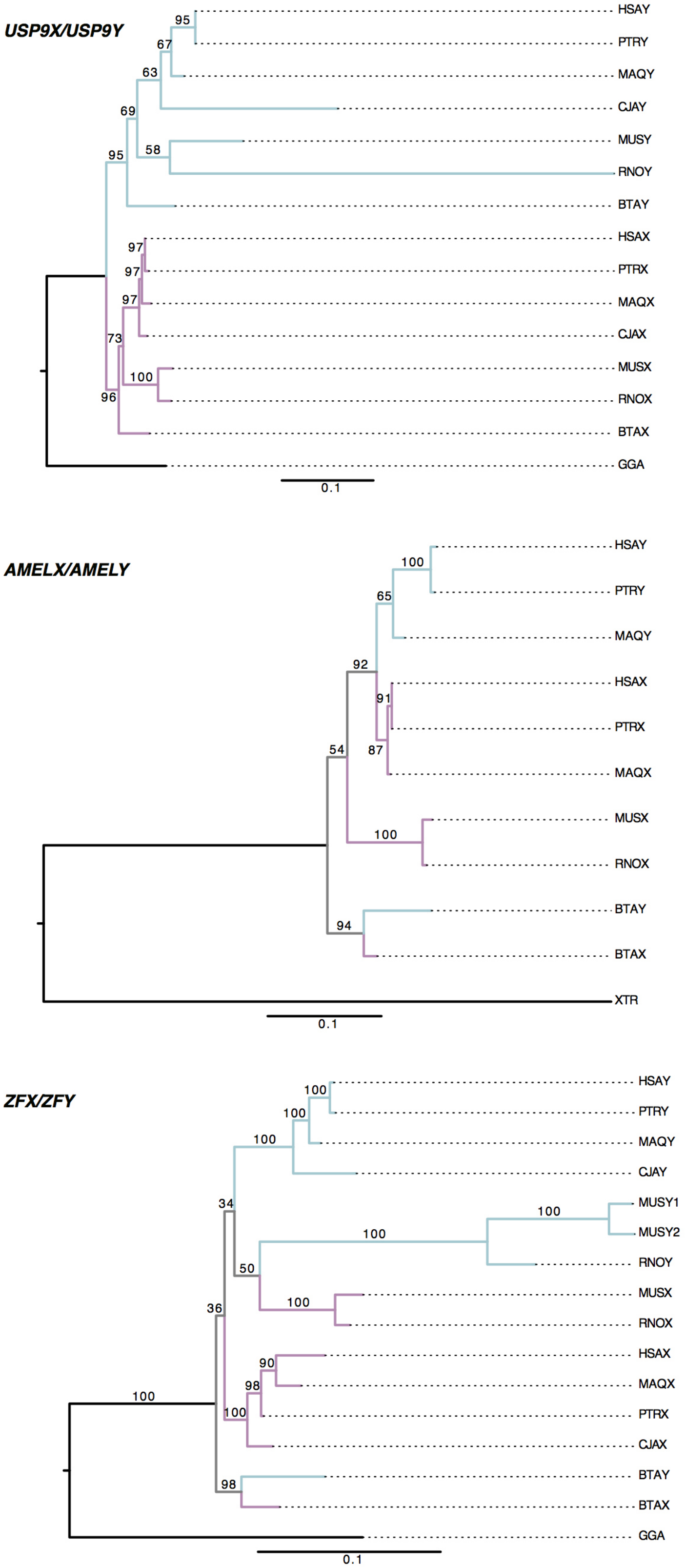
Extended Data Figure 2 | Expression of Y-linked genes across tissues and species. Within each species, the relative expression of each Y-linked gene is shown as a heat map normalized to the male tissue with the highest level of expression of that gene. Expression was calculated from RNA-seq data as reads per kilobase of transcript per million mapped reads. For each gene and species, tissues are arranged in alphabetical order from left to right: brain, cerebellum, colon, heart, kidney, liver, lung, skeletal muscle, spleen and testis. Most single-copy genes (red) are broadly expressed across male tissues, whereas Y-linked genes in multi-copy families (blue) are predominantly or exclusively expressed in testes.



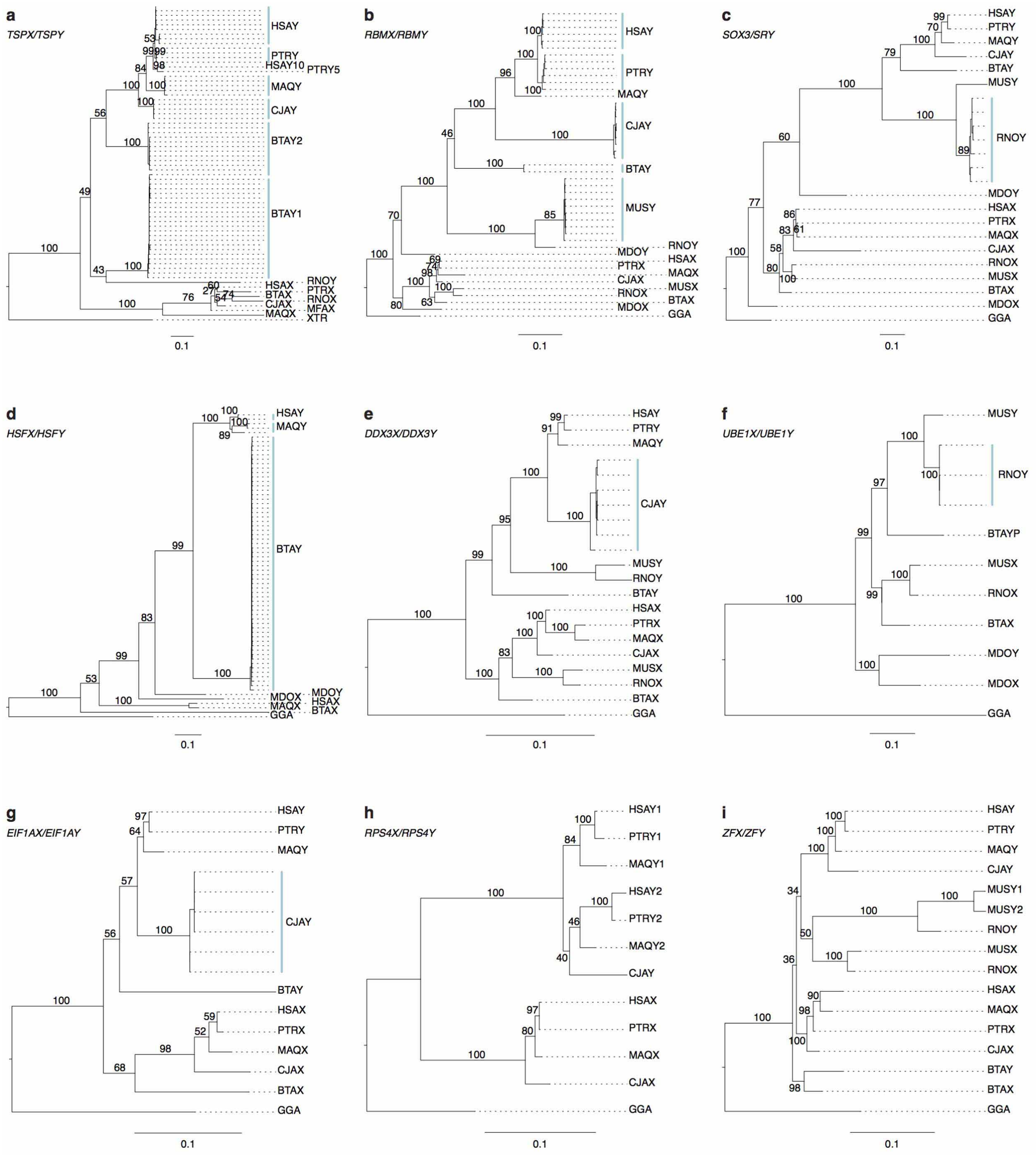
Extended Data Figure 3 | Dot plot of human X orthologues in opossum and chicken. Rectangular dot plots show chromosomal locations of X-orthologous genes in other species. The human X chromosome is composed of a conserved region, orthologous to the opossum X chromosome and a region of chicken chromosome 4, as well as an added region, orthologous to chicken chromosome 1, which has broken in two in the opossum lineage.



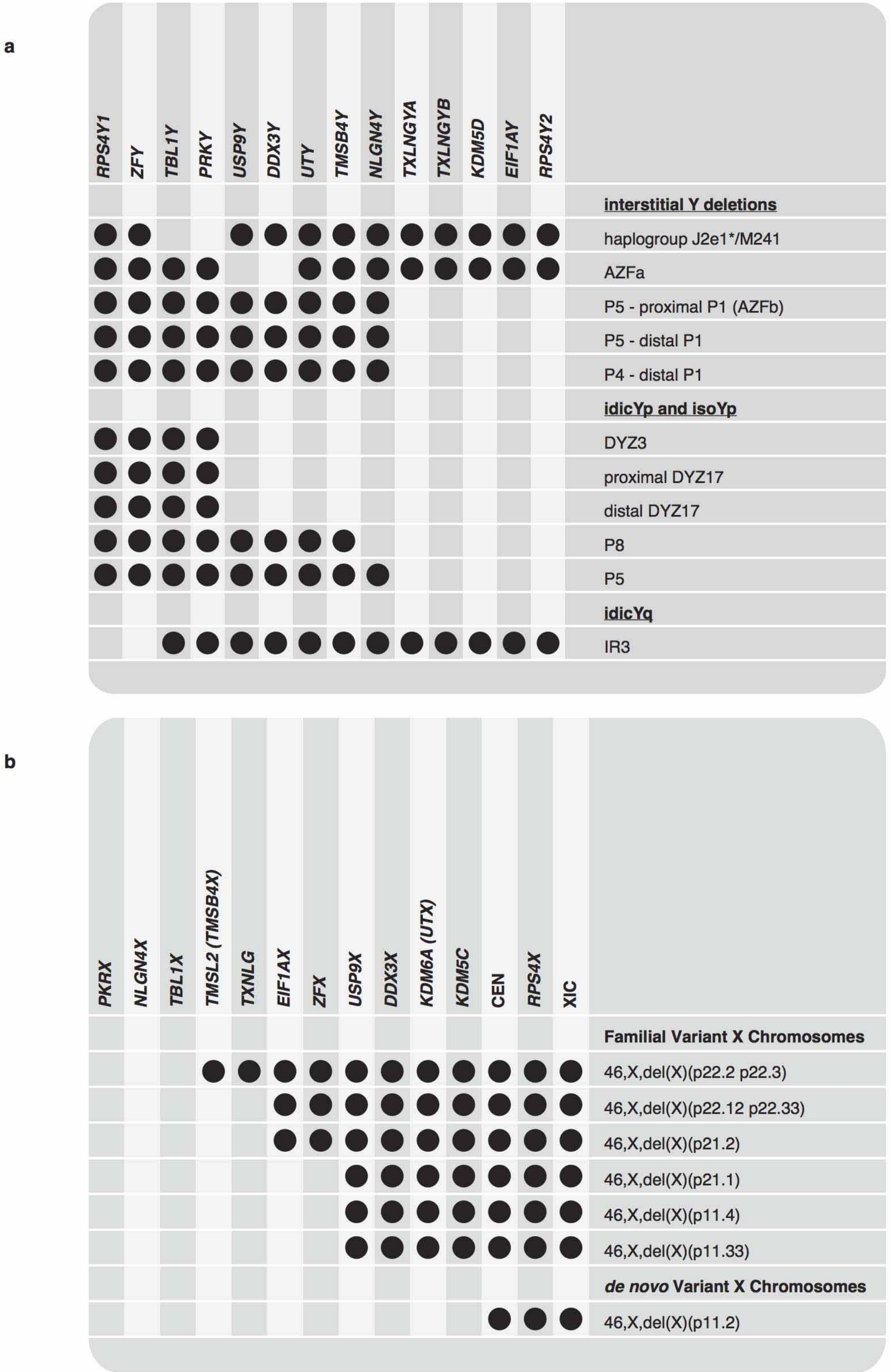
Extended Data Figure 4 | Phylogenetic analysis of stratum one and stratum two genes. Consensus phylogenies reconstructed by DNAML with 100 bootstrap replicates; scale bars represent the expected number of nucleotide substitutions per site along each branch. Phylogenies for ancestral X–Y pair genes from the X-conserved region, shared between placental and marsupial mammals are shown. Adjacent to each tree, pink and light blue bars highlight the positions of the X and Y homologues, respectively; red and dark blue bars highlight the position of placental and marsupial homologues, respectively. Among the three gene pairs from stratum one (*SOX3*/*SRY*, *RBMX*/*RBMY*, and *HSFX*/*HSFY*), Y-linked genes are more closely related to each other than their X-linked orthologues. Among the other gene pairs (*KDM5C*/*KDM5D* and *UBE1X*/*UBE1Y*), marsupial X–Y pairs are more closely related to each other than they are to placental orthologues, suggesting that a second stratum formed independently in the placental and marsupial ineages. Species abreviations: HAS, human; PTR, chimpanzee; MAQ, rhesus; CJA, marmoset; MUS, mouse; RNO, rat; BTA, bull; MDO, opossum; and GGA, chicken.



Extended Data Figure 5 | Phylogenetic tree showing X–Y gene conversion in *AMELX/AMELY* and *ZFX/ZFY*. Consensus phylogenies reconstructed by DNAML with 100 bootstrap replicates; scale bars represent the expected number of nucleotide substitutions per site along each branch. Phylogenies of three ancestral X–Y pair genes from the placental-specific X-added region within stratum 2/3 (*USP9X*/*USP9Y*, *AMELX*/*AMELY* and *ZFX*/*ZFY*) are shown. Within each tree, pink and light blue branches highlight the positions of the X and Y homologues, respectively. *USP9X*/*USP9Y* is a typical stratum 2/3 gene pair; all *USP9Y* genes are more closely related to each other than to any *USP9X* gene. *AMELX*/*AMELY* and *ZFX*/*ZFY* show more complex histories. For example, bull AMELY is more closely related to bull *AMELX* than to any other *AMELY* orthologue. X–Y gene conversion occurred after stratum formation in multiple lineages. Species abbreviations: HAS, human; PTR, chimpanzee; MAQ, rhesus; CJA, marmoset; MUS, mouse; RNO, rat; BTA, bull; MDO, opossum; GGA, chicken; and XTR, *Xenopus tropicalis*.



Extended Data Figure 6 | Y–Y gene conversion within multi-copy gene families. Consensus phylogenies reconstructed by DNAML with 100 bootstrap replicates; scale bars represent the expected number of nucleotide substitutions per site along each branch. Phylogenies for ancestral X–Y pair genes from the X-conserved region, shared between placental and marsupial mammals are shown. Adjacent to each tree, light blue bars highlight the positions of Y-linked genes with high within-species identity and across-species divergence, indicating that gene conversion is more frequent than mutation. **a**-**g**, *TSPY*, *RBMY*, *SRY*, *HSFY*, *DDX3Y*, *UBE1Y* and *EIF1AY* show signs of Y–Y gene conversion; in the species where they are present in multiple copies, they are clustered in arrays of genes. **h**, **i**, *RPS4Y* and *ZFY* do not show signs of recent Y–Y gene conversion; in the species where they are present in two copies, they are dispersed on the Y chromosome. **a**, *TSPY* is present as a multi-copy gene family on the human, chimpanzee, rhesus, marmoset and bull Y chromosomes. Note that 2 distinct families of TSPY emerged in bull. **b**, *RBMY* is present as a multi-copy gene family on the human, chimpanzee, marmoset, mouse and bull Y chromosomes. **c**, *SRY* is present as a multi-copy gene family on the rat Y chromosome. **d**, *HSFY* is present as a multi-copy gene family on the human, rhesus, and bull Y chromosomes. **e**, *DDX3Y* is present as a multi-copy gene family on the marmoset Y chromosome. **f**, *UBE1Y* is present as a multi-copy gene family on the rat Y chromosome. **g**, *EIF1AY* is present as a multi-copy gene family on the marmoset Y chromosome. **h**, *RPS4Y* is present is present as a multi-copy gene family on the human, chimpanzee and rhesus Y chromosomes. *RPS4Y* genes appear to have split into two distinct families before the divergence of primate species, which have not engaged in subsequent gene conversion within each species. **i**, *ZFY* is present as a multi-copy gene family on the mouse Y chromosome. Although *ZFY* participated in multiple independent X–Y gene conversion events after the divergence of placental mammals, there is no evidence of recent Y–Y gene conversion in mouse. Mouse *Zfy1* and *Zfy2* genes are more divergent than human and chimpanzee *ZFY*. Species abbreviations: HAS, human; PTR, chimpanzee; MAQ, rhesus; CJA, marmoset; MUS, mouse; RNO, rat; BTA, bull; MDO, opossum; GGA, chicken; MFA, *Macaca fascicularis*; and XTR, *Xenopus tropicalis*.



a

RPS4Y1	*ZFY*	*TBL1Y*	*PRKY*	*USP9Y*	*DDX3Y*	*UTY*	*TMSB4Y*	*NLGN4Y*	*TXLNGYA*	*TXLNGYB*	*KDM5D*	*EIF1AY*	*RPS4Y2*	
														interstitial Y deletions
●	●			●	●	●	●	●	●	●	●	●	●	haplogroup J2e1*/M241
●	●	●	●			●	●	●	●	●	●	●	●	AZFa
●	●	●	●	●	●	●	●	●						P5 - proximal P1 (AZFb)
●	●	●	●	●	●	●	●	●						P5 - distal P1
●	●	●	●	●	●	●	●	●						P4 - distal P1
														idicYp and isoYp
●	●	●	●											DYZ3
●	●	●	●											proximal DYZ17
●	●	●	●											distal DYZ17
●	●	●	●	●	●	●	●							P8
●	●	●	●	●	●	●	●	●						P5
														idicYq
		●	●	●	●	●	●	●	●	●	●	●	●	IR3

b

PKRX	*NLGN4X*	*TBL1X*	*TMSL2 (TMSB4X)*	*TXNLG*	*EIF1AX*	*ZFX*	*USP9X*	*DDX3X*	*KDM6A (UTX)*	*KDM5C*	CEN	*RPS4X*	XIC	
														Familial Variant X Chromosomes
			●	●	●	●	●	●	●	●	●	●	●	46,X,del(X)(p22.2 p22.3)
					●	●	●	●	●	●	●	●	●	46,X,del(X)(p22.12 p22.33)
					●	●	●	●	●	●	●	●	●	46,X,del(X)(p21.2)
							●	●	●	●	●	●	●	46,X,del(X)(p21.1)
							●	●	●	●	●	●	●	46,X,del(X)(p11.4)
							●	●	●	●	●	●	●	46,X,del(X)(p11.33)
														***de novo* Variant X Chromosomes**
											●	●	●	46,X,del(X)(p11.2)

Extended Data Figure 7 | Viable structually variant sex chromosomes in humans. The presence of the 12 broadly expressed, dosage-sensitive X–Y pair genes and other chromosomal features on structurally variant sex chromosomes are indicated by filled circles. **a**, Viable non-mosaic deletions of X–Y pair genes from the human Y chromosome. The human Y chromosome is susceptible to structural rearrangements due to homology mediated crossing-over between repeated sequences. Crossing-over between tandem repeats creates interstitial deletions, whereas crossing-over in palindrome arms causes the formation of isodicentric chromosomes and isochromosomes. Each Y-linked member of the 12, broadly expressed, dosage-sensitive X–Y gene pairs is deleted in one or more variants, thus no single X–Y pair gene is haplolethal. **b**, Viable deletions of X–Y pair genes from the human X chromosome in females are shown. Reported cases of X chromosome deletions in females are consistent with a collective haplolethality for all 12 broadly expressed, dosage-sensitive X–Y gene pairs in humans. Familial cases, where a variant X chromosome has been transmitted from mother to daughter, are unlikely to be mosaic. The most extensive deletion among familial cases eliminates 7 of 12 genes. The most extensive *de novo* deletion variants eliminate 11 of 12 genes, but mosaicism for 46,XX cells cannot be excluded. No variants remove *RPS4X* because of viability effects mediated by its position between the centromere (CEN) and X-inactivation centre (XIC) on the long arm, rather than haplolethality of *RPS4X* alone.

Extended Data Table 1 | PANTHER statistical over-representation results

PANTHER Pathways	Ancestral_X (554)	XY_Pairs (36)	XY_Pairs (expected)	XY_Pairs (over/under)	XY_Pairs (P-value)
5HT1 type receptor mediated signaling pathway	0	1	0	+	N/A
GABA-B_receptor_II_signaling	0	1	0	+	N/A
Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction	0	1	0	+	N/A
mRNA splicing	0	1	0	+	N/A
Histamine H2 receptor mediated signaling pathway	0	1	0	+	N/A
Enkephalin release	0	1	0	+	N/A
Metabotropic glutamate receptor group I pathway	0	1	0	+	N/A
Metabotropic glutamate receptor group II pathway	0	1	0	+	N/A

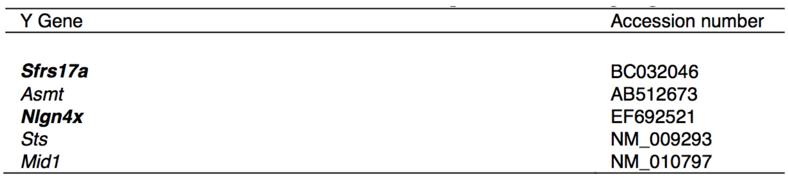
PANTHER GO Biological Process	Ancestral_X (554)	XY_Pairs (36)	XY_Pairs (expected)	XY_Pairs (over/under)	XY_Pairs (P-value)
translation	6	5	0.39	+	7.26E-03
primary metabolic process	240	27	15.6	+	1.97E-02
nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	117	18	7.6	+	2.01E-02
metabolic process	250	27	16.25	+	4.54E-02
protein metabolic process	77	13	5	+	1.23E-01
transcription from RNA polymerase II promoter	71	11	4.61	+	7.22E-01
transcription	71	11	4.61	+	7.22E-01

PANTHER GO Molecular Function	Ancestral_X (554)	XY_Pairs (36)	XY_Pairs (expected)	XY_Pairs (over/under)	XY_Pairs (P-value)
nucleic acid binding	110	21	7.15	+	6.47E-05
translation factor activity, nucleic acid binding	2	3	0.13	+	4.58E-02
binding	200	23	13	+	9.59E-02
translation initiation factor activity	1	2	0.06	+	2.94E-01
structural constituent of ribosome	5	3	0.32	+	6.26E-01
RNA binding	11	4	0.71	+	8.22E-01

PANTHER GO Cellular Component	Ancestral_X (554)	XY_Pairs (36)	XY_Pairs (expected)	XY_Pairs (over/under)	XY_Pairs (P-value)
ribonucleoprotein complex	2	2	0.13	+	2.72E-01

PANTHER Protein Class	Ancestral_X (554)	XY_Pairs (36)	XY_Pairs (expected)	XY_Pairs (over/under)	XY_Pairs (P-value)
RNA binding protein	21	9	1.36	+	1.08E-03
nucleic acid binding	77	16	5	+	1.57E-03
translation initiation factor	1	2	0.06	+	3.55E-01
HMG box transcription factor	1	2	0.06	+	3.55E-01
ribosomal protein	5	3	0.32	+	7.56E-01

Extended Data Table 2 | Accession numbers of mouse pseudoautosomal region genes



Y Gene	Accession number
Sfrs17a	BC032046
Asmt	AB512673
Nlgn4x	EF692521
Sts	NM_009293
Mid1	NM_010797

Extended Data Table 3 | Patients with structurally variant X and Y chromosomes

Deletion Classes	Patients
interstitial Y deletions	
haplogroup J2e1*/M241[61]	
AZFa	WHT2996,WHT3667,WHT4570,WHT4910
P5 - Proximal P1 (AZFb)	WHT4465,WHT3935,WHT4396,WHT4829
P5 – Distal P1	WHT2943,WHT3410,WHT3516,WHT3642, WHT4426,WHT4486,WHT4494,WHT4495
idicYp and isoYp	
DYZ3	WHT2475
proximal DYZ17	WHT2339
distal DYZ17	WHT4596
P8	WHT1876,WHT3062,WHT3954
P5	WHT4818,WHT5060,WHT5139
idicYq and isoYq	
IR3	WHT1301
Familial Variant X Chromosomes	
46,X,del(X)(p22.2 p22.3)[64]	
46,X,del(X)(p22.12 p22.33)[69]	
46,X,del(X)(p21.2)[68,69]	
46,X,del(X)(p21.1)[63,66,67]	
46,X,del(X)(p11.4)[67]	
46,X,del(X)(p11.33)[65,66]	
De novo Variant X Chromosomes	
46,X,del(X)(p11.2)[66,69]	