Supplementary Note 1: SHIMS strategy and error estimates

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:

1. *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: i. high-density filter hybridization using pools of overgo probes, ii. electronic mapping of BAC-end sequences to female genomic sequence, and iii. BAC fingerprint contig analysis. Assembled MSY contigs are verified by PCR using MSY-specific STS markers. Tiling paths of clones are selected for sequencing.

2. *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.

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

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).

The sequencing error rate for the partial MSY sequences for marmoset, mouse, rat, bull, and opossum is approximately one nucleotide per 300 kb.

We ordered and oriented our clone-based contigs using both radiation hybrid mapping and fluorescence *in-situ* hybridization (FISH). 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 Aging 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.

Regions composed of repeats with units less than 30kb 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 Figure 1) are the result of arrays of short, highly identical repeats of this type.

The bull Y-chromosome assembly is interrupted by several large tandem arrays. Two long tandem arrays (estimated size 900Kb and 840Kb) with repeat unit 1.7Kb consist of heterochromatin. These arrays are homologous, but distinguishable. Two gene-containing arrays also interrupt the assembly: a *TSPY1* array with a 7.5Kb repeat unit, and a *PRAME1* array with 21.3Kb unit. All bull 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.

Supplementary Note 2: PANTHER Statistical Overrepresentation Test

We employed the PANTHER statistical overrepresentation test 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 overrepresentation (or underrepresentation) 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 lists 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 overrepresentation we identified was the result of processes that favored 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 homologs (Figure 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 613 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 homolog 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 overrepresented 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.

Supplementary Note 3: Identification and recalibration of evolutionary strata

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]. Orthologs 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 (Figure 1). As expected, none of these pairs have an ortholog on the opossum Y chromosome (Figure 1). Additionally, we note that the opossum orthologs of placental added region genes reside on two autosomes in opossum, chromosome 4 and chromosome 7 (Supplementary Table 2, Extended Data Figure 3). Because the orthologs 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 (Figure 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, while the X-added region contained strata three, four, and five, as well as the freely recombining pseudoautosomal region (PAR)[1,9]. We reexamined 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 (dS) 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 Figure 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 orthologs in the opossum (Figure 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 (dS >= 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 orthologs in chicken as the outgroup. We found that across both placental and marsupial mammals, orthologs of the stratum one genes *SRY*, *RBMY*, and *HSFY* were more closely related to each other than to X-linked homologs (Extended Data Figure 4). Genes from stratum two showed a different pattern; as a group, placental orthologs of *UBE1Y* and *KDM5D* are more closely related to placental X-linked homologs than to their marsupial orthologs (Extended Data Figure 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 (Figure 1, Figure 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 reexamined 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 (Figure 1, Supplementary Table 5). We also excluded *AMELY* and *ZFY*, which participated in interchromosomal gene conversion after stratum formation (Supplementary Table 5, Extended Data Figure 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 (Figure 1, Figure 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 (Figure 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 dataset 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 (Figure 1). The bovine ortholog of *TBL1X* is located in the PAR, and furthermore, *MID1*, which is located between *TBL1X* and *AMELX* on the human X chromosome, has an ortholog straddling the mouse PAR boundary (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 likely the result of lineage-specific X-Y gene conversion events after stratum formation, similar to what has been observed for *ZFY* (Supplementary Table 5, Extended Data Figure 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 MYA for the divergence between old world monkeys and hominoids[12].

NLGN4Y, from stratum four, is present in all four primate species, while *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 (Figure 3)[9,12].

In contrast, human stratum five dates to 32-34 million years ago, prior to 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 (Figure 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 (Figure 1, Supplementary Figure 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 (Figure 1, Supplementary Table 2), demonstrating that, at least while strata are young, genetic decay is both swift and extensive[5,60].

Supplementary Note 4: Modeling kinetics of Y-chromosome decay

We modeled 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 Figure 4.

One-phase exponential decay model:

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

where

$N(t)$ = number of genes at time t

N_0 = number of genes within given stratum in ancestral autosomal/pseudoautosomal portion of genome

K = decay constant

b = baseline (approximated by the number of active ancestral genes within that stratum on human Y chromosome)

Supplementary Note 5: 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 homologs that escape X chromosome inactivation, are collectively haplolethal. Twelve human XY-gene pairs meet this criterion: *RPS4X/RPS4Y, 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-homolog of each X-Y gene pair was deleted in one or more cases (Extended Data Figure 7, 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 homologs 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 Figure 7, 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 center (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 be 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*.

LITERATURE CITED

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. *Nat. Protoc.* **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. U.S.A.* **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. & 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* **140**, 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).