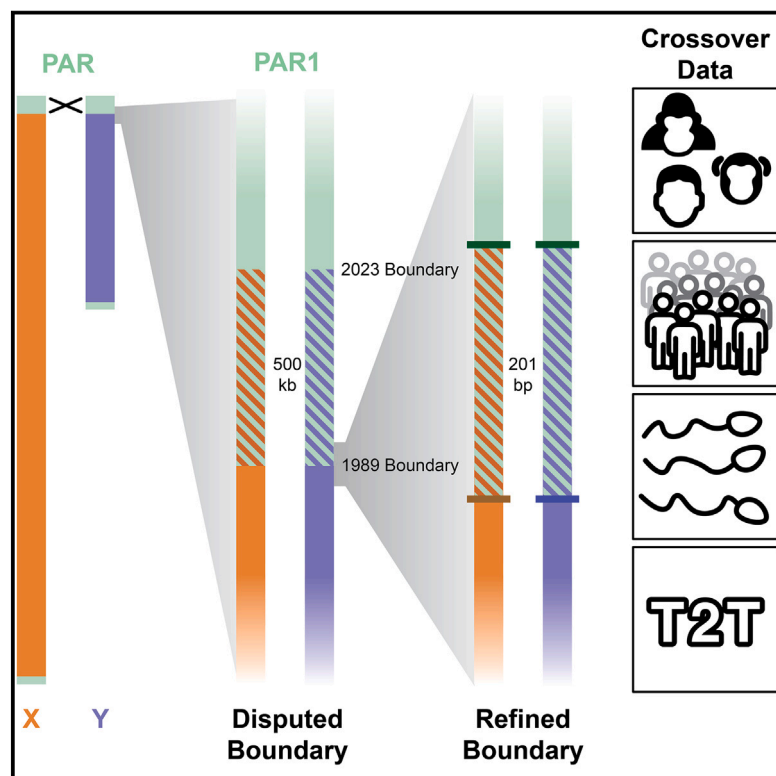


Where is the boundary of the human pseudoautosomal region?

Graphical abstract



Authors

Daniel W. Bellott, Jennifer F. Hughes,
Helen Skaletsky, Erik C. Owen,
David C. Page

Correspondence

dcpage@wi.mit.edu

Controversy has arisen regarding a landmark on the human sex chromosomes—the pseudoautosomal boundary—with recent analyses suggesting a location 500 kb distal to that reported in 1989. Our multifaceted analyses of the available data consistently validate the original position, narrowing the location of the boundary to a 201-bp window.

Where is the boundary of the human pseudoautosomal region?

Daniel W. Bellott,¹ Jennifer F. Hughes,¹ Helen Skaletsky,^{1,2} Erik C. Owen,^{1,3} and David C. Page^{1,2,4,*}

Summary

A recent publication describing the assembly of the Y chromosomes of 43 males was remarkable not only for its ambitious technical scope but also for the startling suggestion that the boundary of the pseudoautosomal region 1 (PAR1), where the human X and Y chromosomes engage in crossing-over during male meiosis, lies 500 kb distal to its previously reported location. Where is the boundary of the human PAR1? We first review the evidence that mapped the PAR boundary, or PAB, before the human genome draft sequence was produced, then examine post-genomic datasets for evidence of crossing-over between the X and Y, and lastly re-examine contiguous sequence assemblies of the PAR-NPY boundary to see whether they support a more distal PAB. We find ample evidence of X-Y crossovers throughout the 500 kb in question, some as close as 246 bp to the previously reported PAB. Our new analyses, combined with previous studies over the past 40 years, provide overwhelming evidence to support the original position and narrow the probable location of the PAB to a 201-bp window.

Introduction

Ninety years ago, Koller and Darlington reported chiasmata between the rat X and Y chromosomes during male meiosis.^{1,2} This led them to propose that the sex chromosomes would each consist of a “pairing segment” where X-Y crossing over occurs and a “differential non-pairing segment.” They proposed that while the differential non-pairing segment was completely sex-linked, the pairing segment should show partial sex linkage, diminishing with genetic (recombinational) distance from the differential segments, and a total crossover frequency approaching 50%.² Nearly 50 years later, Burgoyne coined the term “pseudoautosomal” to describe this partial sex linkage in the pairing segment. We use the terms “pseudoautosomal region” (PAR) to refer to this region of partial sex linkage and non-pseudoautosomal-Y (NPY) and non-pseudoautosomal-X (NPX) to refer to those regions that show complete sex linkage.

Burgoyne posited that to ensure proper chromosome segregation, there was an obligatory crossover between each chromosome pair in male meiosis, and that for the human X and Y, this crossover would occur in a short arm region that was observed to pair under electron microscopy.³ Further, he offered a set of criteria to identify a pseudoautosomal gene: in genetic mapping experiments, it should be partially linked to the X and Y and not to an autosome; it should be physically mapped to the X and Y; and, because genes in the PAR would not require dosage compensation, genes in the PAR should show a strict dependence on sex chromosome dosage, in contrast to NPX genes subject to X chromosome inactivation.³ Only a few years later, in 1986, human *CD99* (formerly known

as *MIC2*) was the first gene demonstrated to fulfill these criteria; it was physically mapped to the X and Y by deletion mapping, was not subject to X inactivation in human-rodent hybrid cells, and was observed to cross over with sex by pedigree analysis.^{4,5}

The human X and Y chromosomes possess two PARs, one at the tip of the short arms and the other at the tip of the long arms (Figure 1A). PAR1, on the short arm (Xp and Yp), is all that remains of a larger ancestral eutherian PAR, while PAR2 is the result of a human-specific transposition from the tip of the long arm of the X chromosome to the Y chromosome⁶ (Figure 1A). The pseudoautosomal boundary (PAB) is the point on Xp and Yp where crossing over between them ceases and the allelic PARs give way to divergent NPX and NPY sequences. The localization of *CD99* provided a foothold for further characterization of PAR1, and in 1989, Ellis and colleagues fine mapped the boundary of PAR1, identifying an Alu repeat sequence inserted on the Y chromosome as the PAB⁷ (Figures 1B and 1C). This identification was supported by complementary genetic and physical mapping techniques. Pedigree-based genetic linkage maps, linkage-disequilibrium-based maps, and single-sperm typing revealed patterns of crossing over between X and Y chromosomes, while somatic cell hybrids, radiation-hybrid mapping, Southern blots, and clone-based sequencing all established physical linkage between the PAR1 and X- or Y-specific DNA sequences.

Hallast and colleagues recently employed a novel approach to locate the PAB of PAR1 by using sequence variation in the human population as a proxy for crossing over.⁸ The recombination rate varies across the human genome, and regions of high recombination are observed to have a higher density of structural variants and indels.

¹Whitehead Institute, Cambridge, MA 02142, USA; ²Howard Hughes Medical Institute, Whitehead Institute, Cambridge, MA 02142, USA; ³Program in Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA; ⁴Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

*Correspondence: dcp@wi.mit.edu
<https://doi.org/10.1016/j.ajhg.2024.09.005>

© 2024 The Author(s). Published by Elsevier Inc. on behalf of American Society of Human Genetics.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

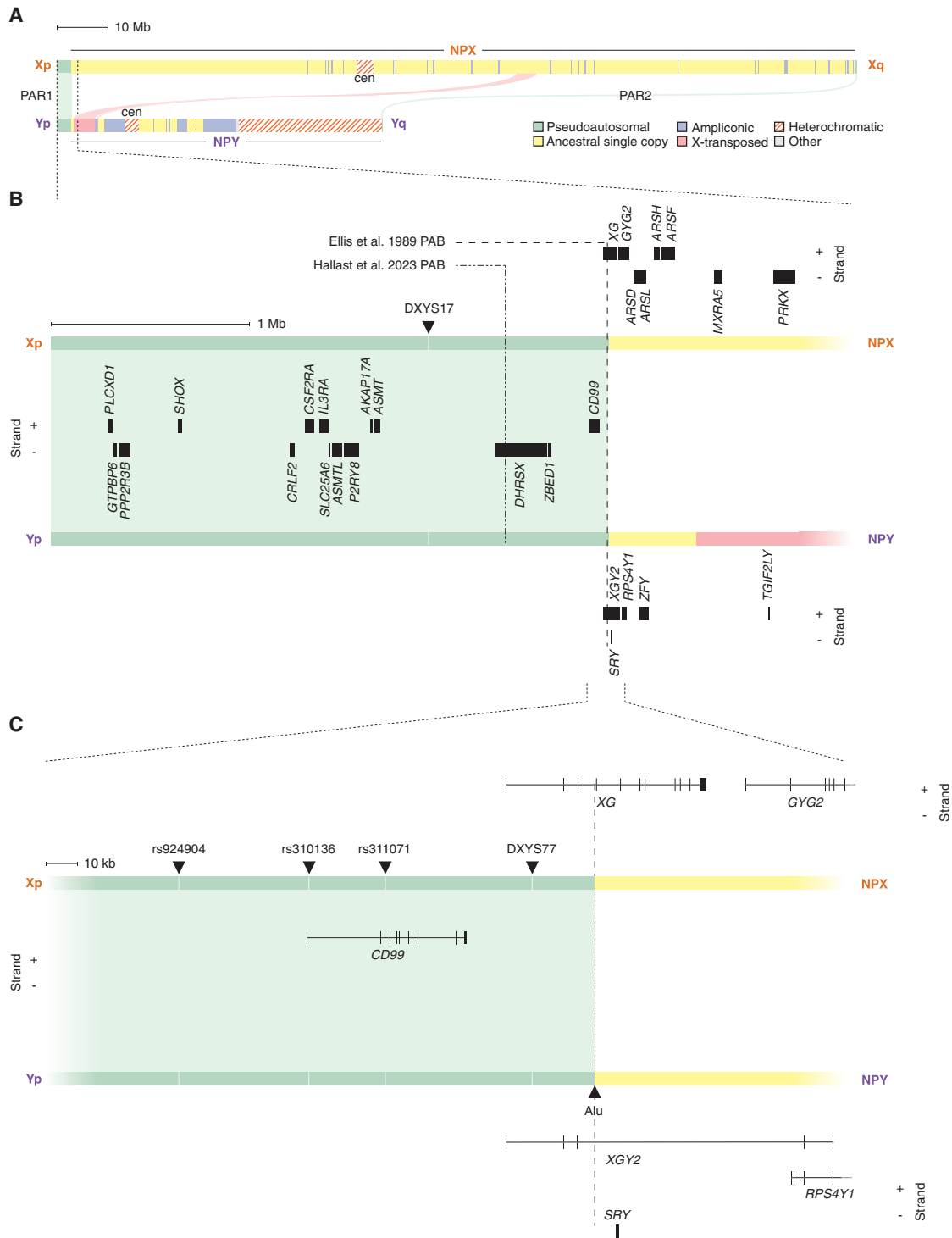


Figure 1. The boundary of human PAR1

(A) Schematic representation of the human X and Y chromosomes, showing the whole of each chromosome, including both PARs. Shading indicates four classes of euchromatic sequence, as well as heterochromatic sequences. Colored arcs indicate regions of >99% X-Y sequence identity. Scale bar, 10 Mb.

(B) Distal 4 Mb of the short arm of each chromosome, enlarged to show PAR1. Dashed lines: 2023 and 1989 PAB. Boxes: genes in PAR1, NPX, and NPY. Arrowheads: genetic markers. Scale bar, 1 Mb.

(C) 250 kb of each chromosome near the 1989 PAB. Genes shown as lines with arrowheads indicating direction of transcription, exons as boxes. XG and XGY2 share three exons in the PAR. Scale bar, 10 kb.

Both recombination and variation are highest in PAR1, followed by the autosomes, the NPX, and are lowest in the NPY. Hallast and colleagues observed that the 500-kb

segment immediately distal to the 1989 PAB shows a low density of structural variants and indels among their sequenced Y chromosomes, and they reasoned that this

observation is best explained by reduced or absent recombination in this interval. Hallast and colleagues concluded that “our data suggest that the boundary between the recombining pseudoautosomal region 1 and the non-recombining portions of the X and Y chromosomes lies 500 kb away from the currently established boundary.”

This is indeed an arresting discovery—is it possible that the location of this critical landmark on the human sex chromosomes was mistaken, and the size of PAR1 has been overestimated by 20% for 35 years? What is the nature of the evidence for the 1989 PAB? First, we will review evidence accumulated over the past 40 years, and then we will present new analyses taking advantage of expanded genomic datasets, which all converge on confirming, and slightly refining, the original position of the PAB.

Establishing the PAB

The earliest genetic maps of PAR1 were based on analysis of human pedigrees using cloned DNA sequences as probes to identify restriction fragment length polymorphisms on Southern blots.^{5,9–13} Rouyer and colleagues genotyped three pseudoautosomal markers across eight Centre d’Etude du Polymorphisme Humain (CEPH) families to produce the first genetic linkage map of PAR1.¹¹ The map revealed a gradient of sex linkage from the most proximal to the most telomeric marker and covered a distance of 50 cM (centimorgans), indicating that, on average, one crossover occurs in PAR1 during each male meiosis.¹¹ However, *DXYS17*, the marker most closely linked to sex, was still 14 cM distal to the PAB (Figure 1B). Subsequently, Goodfellow and colleagues re-mapped these markers, along with *CD99*, across an independent set of three pedigrees.⁵ *CD99* was the proximal marker, mapping only 2.2 cM distal to biological sex.⁵ Thus, the PAB for PAR1 must lie between *CD99* and the sex-determining locus.

The tight linkage of *CD99* with sex inspired researchers to attempt to isolate the boundary between PAR1 and the NPY using *CD99* as a probe. Using a 49,XYYYY cell line as a highly enriched source of Y-linked DNA, Pritchard and colleagues used rare-cutting restriction enzymes, both singly and in pairs, to generate long restriction fragments, which they resolved by pulsed-field gel electrophoresis.¹⁴ They identified a 200-kb restriction fragment that hybridized to both *CD99* and NPY sequences.¹⁴ Subsequently, Ellis and colleagues further refined this interval, beginning with a 110-kb restriction fragment containing their most proximal *CD99* and most distal NPY probes and gradually narrowed this down to a 0.8-kb restriction fragment containing the boundary between PAR1 and NPY.⁷ Armed with precise knowledge of the most proximal PAR sequences, they also isolated a 0.9-kb restriction fragment containing the boundary between PAR1 and NPX.⁷ They sequenced both fragments and found they were 99% identical distal to a 303-bp Alu repeat insertion in the Y-derived sequence.⁷ Proximal to the Alu insertion, X-Y identity drops to 77% for another 245 bases and fails

to align further.⁷ They reasoned that the high X-Y identity in the sequence distal to the Alu insertion was maintained by crossing over and that the distal end of the Alu insertion marked the PAB (Figure 1C).

To establish that this sequence did constitute the PAB, it was critical to find additional evidence of crossing over. Ellis and colleagues took a two-pronged approach, examining homologous X and Y sequences across catarrhine primates and sequencing this region from diverse human populations to identify polymorphisms.^{15,16} Their evolutionary analysis found that while the Alu insertion was found across the Y chromosomes of great apes, it was absent from Old World monkeys.¹⁵ However, even Old World monkeys maintained a crisp drop off in X-Y identity between regions distal and proximal to the Alu insertion site.¹⁵ This suggested that X-Y identity in the distal region was maintained by crossing over since the split of Old World monkeys and great apes. By surveying 57 Y chromosomes and 60 X chromosomes from ten human populations, Ellis and colleagues identified single-nucleotide polymorphisms (SNPs) to infer patterns of crossing over.¹⁶ While there were fixed X-Y differences at positions 41 and 45 bases distal to the Alu insertion, they observed a site with a polymorphism shared by the X and the Y only 274 bases distal to the Alu.¹⁶ Between these two sites, all Y chromosomes shared SNPs with a common X chromosome haplotype. This provided strong evidence that crossovers had occurred between 45 and 274 bases distal to the Alu insertion, localizing the human PAB to this 229-base interval.

Following the localization of the 1989 PAB, crossovers have repeatedly been observed in the 500 kb distal to this site by independent methods. Pedigree-based genetic maps rely on crossovers to infer the linear order and genetic distances among loci along a chromosome. In contrast, single-sperm typing isolates and genotypes single haploid sperm cells to reconstruct haplotypes without pedigree analysis and infer meiotic recombination events. Both methods have observed crossing over between the 1989 PAB and markers only tens of kilobases away.

In producing the second iteration of the Rutgers Map of the human genome, Matisse and colleagues mapped 28,121 markers on CEPH pedigrees, including three markers within *CD99* (rs924904, rs310136, and rs311071)¹⁷ (Figure 1C). These markers mapped from 5.12 to 2.79 cM distal to the 1989 PAB.¹⁷ Flaquer and colleagues added more pseudoautosomal markers to the third iteration of the Rutgers map and mapped two of these markers even closer to the PAB: rs310136 87.6 kb and 0.7 cM away and rs311071 66.3 kb and 0.0007 cM distal to the PAB.¹⁸ Subsequently, Hinch and colleagues generated a fine-scale map of the PAR; they observed 126 paternal crossovers in PAR1, of which at least 14 were in the 500 kb immediately distal to the 1989 PAB.¹⁹ This region covers 5.1 cM in their map and includes a recombination hotspot within *XG*, which straddles the PAB on the X chromosome but is disrupted on the Y chromosome.¹⁹

Schmitt and colleagues typed 903 flow-sorted sperm from two donors to map recombination among five loci in PAR1 and used 2,005 sperm from a third donor to map *DXYS77*, the marker closest to the 1989 PAB, at higher resolution.²⁰ *DXYS77* is located in an intron of *XG*, 19.6 kb and 0.77 cM distal to the PAB (Figure 1C). Subsequent single-sperm typing analyses have not mapped PAR1 at such high resolution but have also observed crossovers in the 500 kb distal to the 1989 PAB.^{21,22}

In summary, several lines of evidence that revealed both historical and contemporary crossovers between X and Y chromosomes were used to support the 1989 PAB. Here, we take advantage of a variety of publicly available, post-genomic datasets to re-examine this question in light of the more distal PAB proposed by Hallast and colleagues.⁸ We examine crossover events during primate evolution by aligning X and Y chromosomes from ape and Old World monkey genome assemblies. Similarly, we use sex differences in minor allele frequency in a genome-wide catalog of common human genetic variation to reveal historical patterns of crossing over in the human population. Analysis of several independent single-sperm sequencing datasets allowed us to locate ongoing crossovers in the PAR of living individuals. Lastly, we use the known phylogenetic relationship between the NPY sequences in the genomes assembled by Hallast and colleagues⁸ to look for patterns of nucleotide substitution that are best explained by crossing-over. We find that all of these lines of evidence support the 1989 PAB.⁷

Methods

Primate X-Y alignments

We downloaded primate X- and Y-linked sequences spanning the PAB from GenBank (Table S1). Prior to alignment, we removed the Alu insertion at the 1989 PAB from ape Y-linked sequences, as well as an independent Alu insertion 300 bp proximal to the 1989 PAB in the colobus NPY. We used PRANK (v.170427)²³ to generate a multiple alignment of sequences distal and proximal to the 1989 PAB (Data S1). We used various windows of these multiple alignments to infer the phylogenetic relationships between sequences with PhyML (v.3.3.20190909)²⁴ with 100 bootstrap replicates.

Human polymorphisms from 1000 Genomes

We downloaded phased individual-level variants from the PAR from the 1000 Genomes Project.²⁵ For each of 118,327 polymorphic sites in PAR1, we tabulated the number of males and females who were homozygous for the reference allele, heterozygous, and homozygous for the alternative allele (Table S2). We used these counts to calculate sex-specific minor allele frequencies for males and females.

Single-sperm sequencing

We obtained raw sequence data Database of Genotypes and Phenotypes (dbGaP) study dbGaP: phs001887 from the NCBI Sequence Read Archive. Using Bowtie2 (version 2.4.2),²⁶ we aligned reads to a version of hg38 (GenBank: GCF_000001405.40) with the chrY PAR masked. For each donor, we combined PAR alignments from

all haploid cells using samtools (version 1.11),²⁷ marked duplicate reads with sambamba (version 0.8.2),²⁸ and called variants at sites in dbSNP (<https://www.ncbi.nlm.nih.gov/snp>) with freebayes (version 1.3.2).²⁹ We used vcfFilter³⁰ to restrict downstream analyses to bi-allelic SNPs with $Q > 20$, more than two reads each of reference and alternate alleles, and allele frequencies between 0.25 and 0.75. We identified X- and Y-bearing sperm by the depth of reads mapping to NPX and NPY sequences. In our analyses of the sperm-seq data from phs001887, we also excluded reads with cell barcodes associated with cell doublets and bead doublets by Bell and colleagues.³¹

For each donor, we reconstructed X and Y haplotypes across PAR1 in two passes. In each pass, we assigned the alleles at each variant site to either the X or Y donor haplotype to minimize the number of crossovers with more proximal sites, then within each sample, localized crossovers between sites with variants assigned to different donor haplotypes. In the first pass, we flagged markers in tight double recombinants (closer than 278,148 bases; that is, one-tenth the length of PAR1, or ~ 5 cM) as unreliable and identified non-crossover samples. In the second pass, we excluded unreliable markers flagged in the first pass, as well as any markers with more than 10% discordant genotypes in non-crossover samples. We then removed any remaining tight double recombinants before calculating the local recombination rate.

Reanalysis of Y chromosomes sequenced by Hallast and colleagues

Hallast and colleagues produced contiguous assemblies of the PAR-NPY boundary from ten samples (HG01890, HC02666, HC19384, NA19331, HG02011, HG03371, NA19317, HC18989, HG03009, and HG00358)⁸; we downloaded their assemblies and extracted the PAR contigs for further analysis.

To analyze clock-like evolution on the Y chromosome, we first masked simple repeats with RepeatMasker (version 4.1.5) and tandem repeats with trf (version 4.09),³² then aligned all ten Y chromosome assemblies with progressiveMauve (build date May 15, 2023 at 08:46:41).³³ We extracted 20-kb windows of the multiple alignment, excluding columns of gaps and masked characters, at 1-kb intervals. We used tree-puzzle (version 5.2)³⁴ to conduct a likelihood ratio test (LRT) to compare a model where sequences evolve using a single clock-like rate based on their relationships from the Y chromosome haplogroup tree vs. a model where rates are allowed to vary across branches. We obtained p values from the test statistic reported by tree-puzzle (Table S4) using the pchisq function and adjusted for multiple testing by Bonferroni correction with the p.adjust function in R (version 4.1.1).³⁵

To look for polymorphisms shared among X and Y chromosomes, we aligned homologous X and Y sequences from the ten genomes with contiguous PAR sequence and RP11-400O10 (GenBank: AC006040.3), a human Y chromosome bacterial artificial chromosome (BAC) clone spanning the 1989 PAB, and manually edited the resulting alignments in Unipro UGENE (version 49.1)³⁶ (Data S3).

Results

Primate X-Y alignments

We took advantage of the rich genomic resources from multiple species to revisit the evolutionary analyses of Ellis and colleagues⁷ using high-quality assemblies of primate X and Y chromosomes. Within the PAR, crossing-over will

result in the DNA sequences of Y chromosomes being more similar to X chromosomes of the same species than to Y chromosomes of other species. Conversely, in the NPY, the absence of crossing-over will result in NPY sequences from different species being more similar to each other than to NPX sequences from the same species. We curated high-quality sequences adjacent to the 1989 PAB from the NPX and NPY from seven catarrhine primate species: human, chimpanzee, gorilla, orangutan, siamang, rhesus macaque, and colobus monkey (Table S1; Data S1). We constructed a multiple alignment from these sequences and used these alignments to construct trees using a range of window sizes from the alignment, both distal and proximal to the 1989 PAB (Figure 2). Across all window sizes, even as narrow as 1 kb from the 1989 PAB (Figure 2D), we observe that sequences distal to the 1989 PAB group together by species, while sequences proximal to the 1989 PAB group together by chromosome. Examining this multiple alignment more closely, we find human X and Y chromosomes sharing an A>G substitution only 93 bases distal to the 1989 PAB (Figure 3). Since this substitution is also present on the chimpanzee Y, this suggests that it originated on the Y chromosome in the last common ancestor of humans and chimpanzees and crossed over to the X chromosome in humans (Figure 3). The Alu element, from the catarrhine-specific AluY family,³⁷ was inserted into the Y chromosome after apes diverged from Old World monkeys and has not crossed over to the X. This Alu insertion must have occurred at or near the ancestral catarrhine PAB; only 11 bases proximal to this site, all X chromosomes share a T where all Y chromosomes carry a G or A (Figure 3), confirming that crossing over was suppressed in this proximal region before the divergence of apes and Old World monkeys.

Human polymorphisms from 1000 Genomes

To gain additional insight into the history of crossing-over within the human population, we analyzed polymorphism data from the 1000 Genomes Project.²⁵ Strictly Y-linked regions with low or absent crossing over with the X chromosome will accumulate mutations that are male specific, leading to sex biases in the frequency of minor alleles. Using data from 1,599 males and 1,603 females, we calculated sex-specific minor allele frequencies for 118,327 polymorphic sites across PAR1 (Table S2). We plotted the sex difference in minor allele frequency throughout PAR1 and observed that the differences between male and female minor allele frequencies are typically small, even in the 500 kb immediately distal to the 1989 PAB (Figure 4A). Within 10 kb of the 1989 PAB, alleles are more sex biased, and this sex bias in minor allele frequencies increases rapidly as one moves proximally, nearing the maximum possible value of 0.5 (Figure 4A). Distal to this 10-kb segment adjacent to the 1989 PAB, male and female minor allele frequencies closely mirror each other, indicating that crossing over has intermixed and homogenized the sets of alleles on the X and Y chromosomes (Figure 4B).

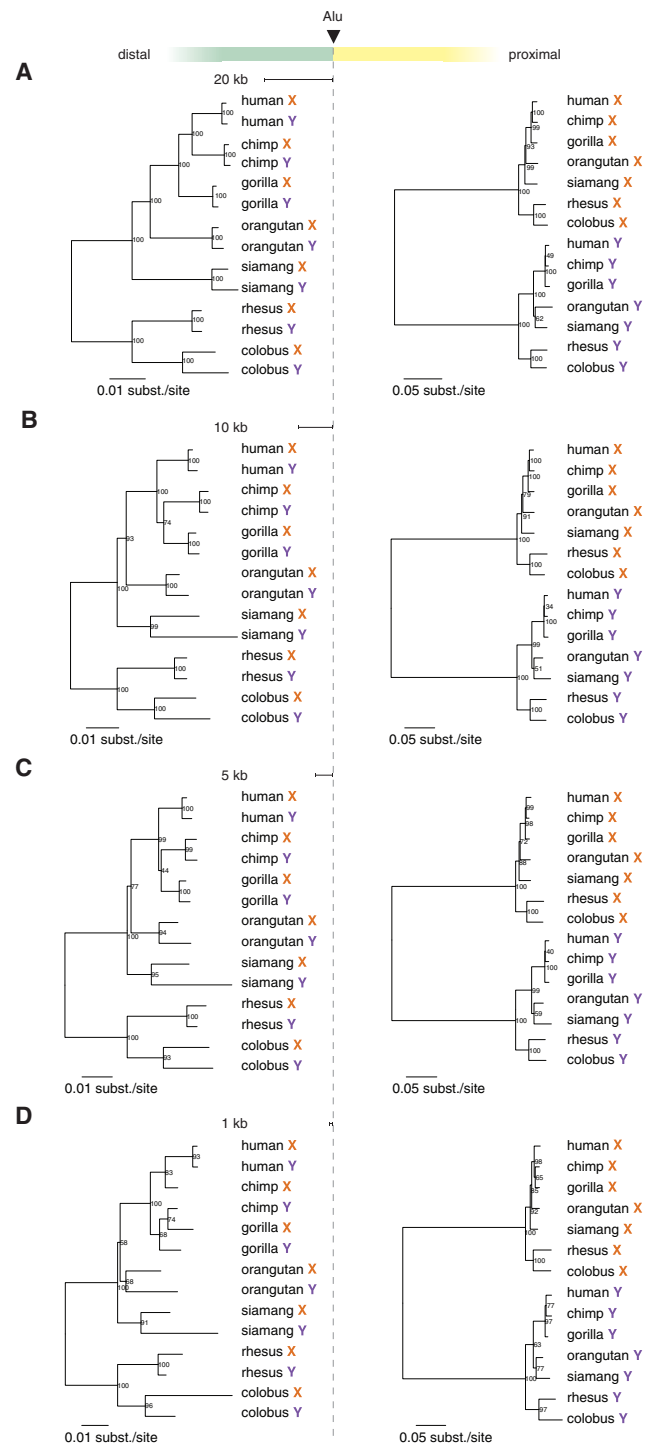


Figure 2. Phylogenetic relationships between primate X and Y sequences differ distal and proximal to 1989 PAB

Maximum likelihood trees based on multiple alignment of primate sex chromosome sequences in windows (A) 20 kb, (B) 10 kb, (C) 5 kb, or (D) 1 kb upstream and downstream of the 1989 PAB; nodes show support for branching order from 100 bootstrap replicates. Scale bars, branch length in substitutions per site.

Single-sperm sequencing

To capture ongoing crossover events in PAR1 from living individuals, we analyzed a massive single-sperm sequencing dataset (dbGaP: phs001887) containing

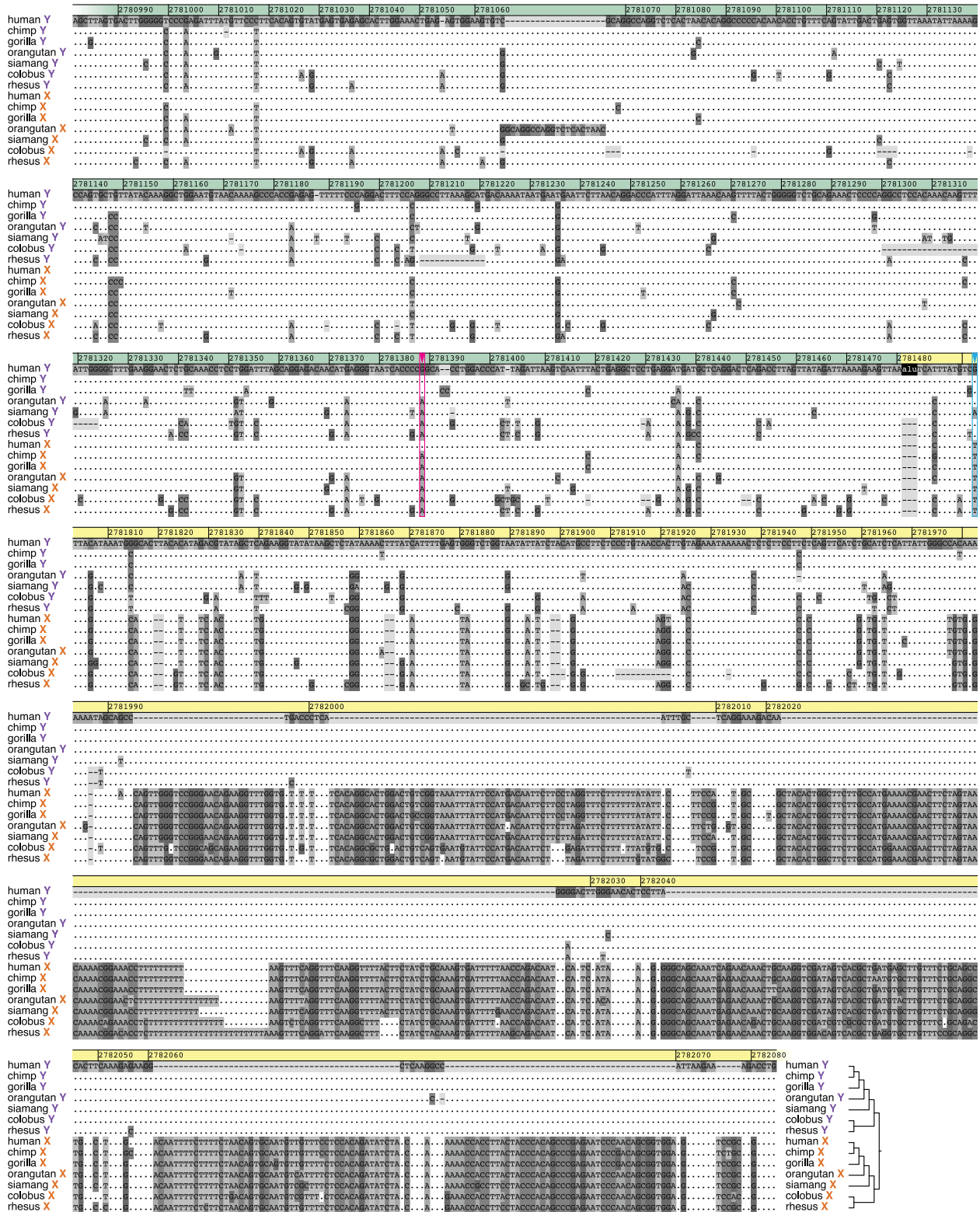


Figure 3. Multiple alignment of primate sex chromosomes shows sharp change in X-Y identity at the 1989 PAB
 Multiple alignment of primate sex chromosome sequences showing the 1989 PAB at or near the ape-Y-specific Alu insertion (black). Dots indicate identity with human Y sequence; coordinates relative to hg38 reference. Tree indicates phylogenetic relationship in NPX and NPY regions. Magenta arrowhead and box marks crossover from Y to X in the human lineage; cyan arrowhead and box a fixed NPX/NPY difference predating the Alu insertion in apes.

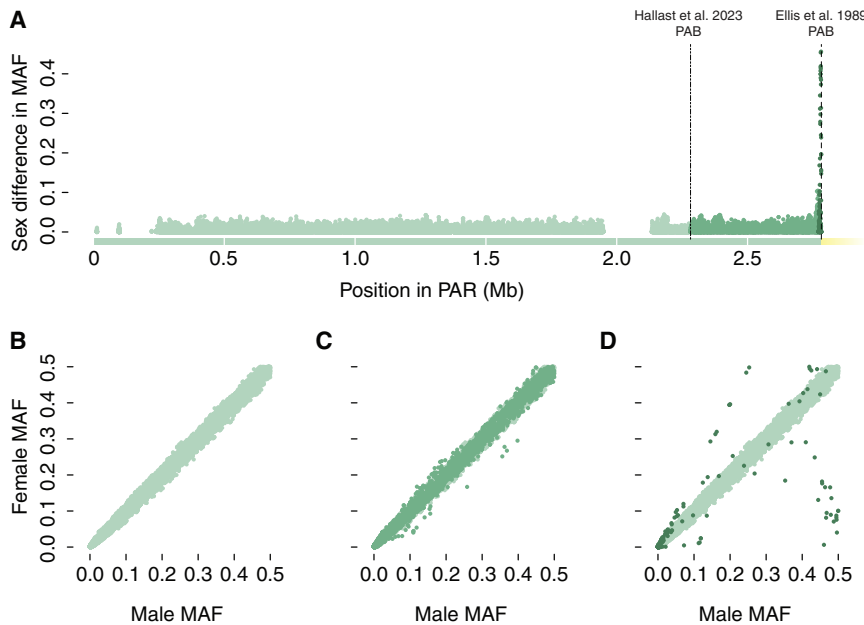


Figure 4. Sex differences in MAF only occur within 10 kb of the 1989 PAB

Minor allele frequency (MAF) calculated from 1000 Genomes data for the human PAR. Dark green, sites <10 kb from the 1989 PAB; medium green, <500 kb from the 1989 PAB; light green, all other sites.

(A) Absolute sex difference in MAF plotted against hg38 location. 2023 and 1989 PAB indicated by dashed lines.

(B–D) Male vs. female MAF; (B) tight correlation between male and female MAFs distal to the 2023 PAB; (C) similar correlation at most positions between the 2023 and 1989 boundaries; except (D) in the 10 kb immediately distal to the 1989 PAB.

PAR can be transferred from Y chromosomes to X chromosomes, or from one Y chromosome haplogroup to another through an X chromosome intermediate, across generations. As a result, for sequences within the PAR,

Illumina reads from 31,228 sperm across 20 donors, which has not been previously analyzed for PAR1 crossovers.³¹ We identified X- and Y-bearing sperm from the depth of reads aligning to the NPX and NPY and used this information to phase variant sites in PAR1. Across all donors, we observed 795 sperm with unambiguous crossovers between the 1989 and 2023 PABs (Table S3). We calculated the recombination rate between all genotyped variants in PAR1 (Figures 5A and S1) and estimate the total genetic distance between the 1989 PAB and 2023 PAB to be 4.78 cM (Figure 5B). We observe a local increase in the recombination rate close to the 1989 PAB (Figure 5A), parallel to what Hinch and colleagues observed using a pedigree-based genetic map.¹⁹ We note that this local pattern of increased recombination near the 1989 PAB is not present across all donors (Figure S1) consistent with the observation that PAR1 hotspots are rapidly evolving and differ among human populations.¹⁹

Reanalysis of Y chromosomes sequenced by Hallast and colleagues

All three of these fully independent analyses confirmed the location of the 1989 PAB, so we asked whether the sequences generated by Hallast and colleagues⁸ might also provide evidence for crossing over in the 500 kb distal to the 1989 PAB. Of the 43 individual human male genomes assembled by Hallast and colleagues, ten have contiguous PAR1 assemblies, supported by high-quality differences between long reads originating from the X and Y chromosomes.⁸ We restricted our analyses to these ten genomes where the provenance of the sequences within the PAR are unambiguous.

During human evolution, substitutions are predicted to accumulate on the NPY in a clock-like fashion with no recombination. However, substitutions that arise in the

branch lengths calculated using the NPY haplogroup tree will show deviations from clock-like behavior. We constructed a multiple alignment of all ten contiguous Y-linked PARs (Data S2). Using the phylogenetic relationships between these sequences based on their NPY haplogroup,⁸ we conducted an LRT in 20-kb sliding windows of the alignment, comparing a single-clock model where all branches accumulate substitutions at the same rate vs. a local-clock alternative where rates are allowed to vary among branches (Figure 6; Table S4). We observe a crisp change in the results of this test that coincides with the 1989 PAB (Figure 6). Windows proximal to the 1989 PAB are almost universally consistent with clock-like evolution along the haplogroup tree, as expected for NPY sequences. The rare exceptions may represent X-Y gene conversion events in homologous sequences. Distal to the 1989 PAB, the single-clock model can be rejected in the majority of windows. This can readily be explained by a history of crossing over distal to the 1989 PAB so that the Y-haplogroup tree does not accurately describe the relationship between these sequences.

Encouraged by these results, we further examined the homologous X and Y sequences from these ten genomes to ask how close to the 1989 PAB we could observe evidence of crossing-over. Following the procedure of Ellis and colleagues,⁷ we constructed a multiple alignment of these ten diverse X- and Y-linked sequences along with sequence from a BAC clone from the Y chromosome spanning the 1989 PAB (Figure 7; Data S3) and looked for polymorphisms shared by X and Y chromosomes. We observe the same fixed differences between human X and Y chromosomes at positions 41 and 45 bases distal to the Alu insertion that Ellis and colleagues described more than 30 years ago (Figure 7). Similarly, we see that all Y chromosomes share a haplotype commonly found on X

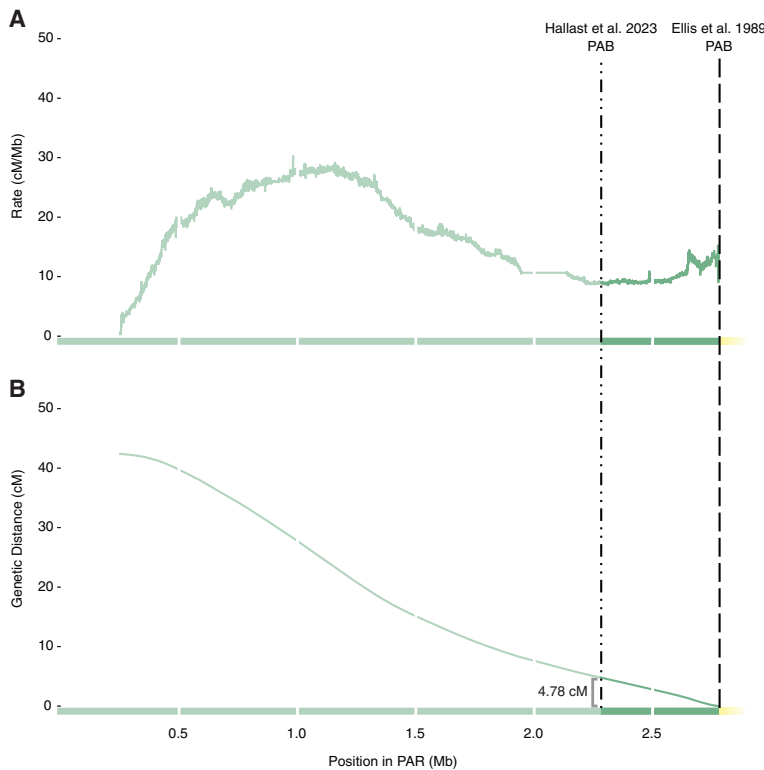


Figure 5. Single-sperm sequencing reveals cross-overs between 2023 and 1989 PABs

Summary of sperm-seq results across 20 donors showing (A) average local recombination rate between genotyped markers within PAR1 and (B) total genetic distance from the 1989 PAB. Medium green, region between 2023 PAB and 1989 PAB; light green, all other sites; white tick marks every 500 kb.

chromosomes up to 245 bases distal to the Alu insertion (Figure 7). We observe five previously reported polymorphic sites (rs147962110, rs189310269, rs180883975, rs2534635, and rs5982851) where reference and alternate alleles are shared by X and Y chromosomes (Figure 7). The closest site, rs5982851, is only 246 bases distal to the Alu insertion that marks the PAB (Figure 7) while the next closest, rs2534635, was previously observed by Ellis and colleagues.¹⁶ Thus, rather than overturning the 1989 PAB, the sequences produced by Hallast and colleagues⁸ reveal a historical recombination event 28 bases closer to the 1989 PAB than Ellis and colleagues⁷ initially observed, narrowing the possible location of the PAB to the 201-base interval between rs5982851 and the first fixed difference between the X and Y chromosomes.

Discussion

Our present understanding of the biology of the sex chromosomes has been informed by close study of crossing over between the X and Y chromosomes. During meiosis, crossing over generates a physical link between each pair of homologous chromosomes, providing the tension required to ensure that homologs can be segregated faithfully. The PARs are responsible for providing this obligate crossover between the X and Y chromosomes during male meiosis. Without this crossover, meiosis generates either nullisomic sperm that produce inviable non-mosaic 45,X embryos³⁸ or XY sperm that produce 47,XXY offspring that develop Klinefelter syndrome. In contrast,

when there are crossovers involving the NPX or NPY, sperm will carry structurally anomalous X and Y chromosomes that produce 46,XX male (disorder of sex development [DSD]) and 46,XY female (DSD) offspring. All these meiotic anomalies result in offspring that are inviable or infertile, imposing strong selection to maintain the PAR. The highly differentiated human X and Y chromosomes evolved from what were once an ordinary pair of autosomes that crossed over freely. This divergence was the result of a series of rearrangements over the last 200 million years that first suppressed crossing over around *SRY*, the testis-determining gene, initially creating a PAB, and then repeatedly shifted

the PAB to expand the NPY at the expense of the PAR.³⁹ The absence of crossing over in males radically altered both the NPX and NPY; 97% of ancestral NPY genes were lost to genetic decay, while their counterparts on the NPX evolved new regulatory mechanisms to maintain the ancestral gene dose in both sexes.⁴⁰ As a result, the PAB forms the frontier between the sex-specific biology of the NPX and NPY, and the sex-shared biology of the PAR. As we continue to investigate the genetics of sex differences, the PAB remains a critical landmark on the human sex chromosomes.

For 40 years the PAB has been defined by the pattern of crossing over. The available evidence overwhelmingly supports the 1989 PAB in PAR1, rather than a site 500 kb more distal. Crossing over between the X and Y chromosomes has maintained high X-Y identity across the region distal to the 1989 PAB since the divergence of apes and Old World monkeys around 30 million years ago.⁴¹ Within the human population, crossovers occurred as close as 246 bases to the 1989 boundary, as inferred from alleles at polymorphic sites shared by both X and Y chromosomes. Human pedigrees and single-sperm typing data record ongoing crossovers throughout the PAR, including the 500 kb adjacent to the 1989 boundary. It is therefore clear that the absence of structural variation cannot be used as a proxy for the absence of crossing over. However, we note that our findings do not otherwise challenge the methods or findings of Hallast and colleagues,⁸ which in other respects agree closely with previous characterizations of the human Y chromosome.



Figure 6. Y chromosome sequences distal to the 1989 PAB are inconsistent with phylogenetic tree of human NPY haplotypes
Results of LRT comparing single-clock vs. local-clock models of evolution on the NPY haplogroup trees in 20-kb sliding windows across multiple alignment of ten contiguous Y chromosome sequences produced by Hallast and colleagues.⁸ Bonferroni corrected p values for LRT plotted vs. hg38 position. 2023 and 1989 PAB indicated by dashed lines. Windows distal to the 1989 PAB repeatedly reject the single-clock model.

Data and code availability

This study did not generate datasets. Code for analysis of sperm-seq data is available at: https://github.com/dwbellott/sperm_seq_scripts.

Acknowledgments

This work was supported by NIH Pre-Doctoral Training Grant T32GM087237, the Whitehead Institute, the Howard Hughes Medical Institute, the Simons Foundation Autism Research

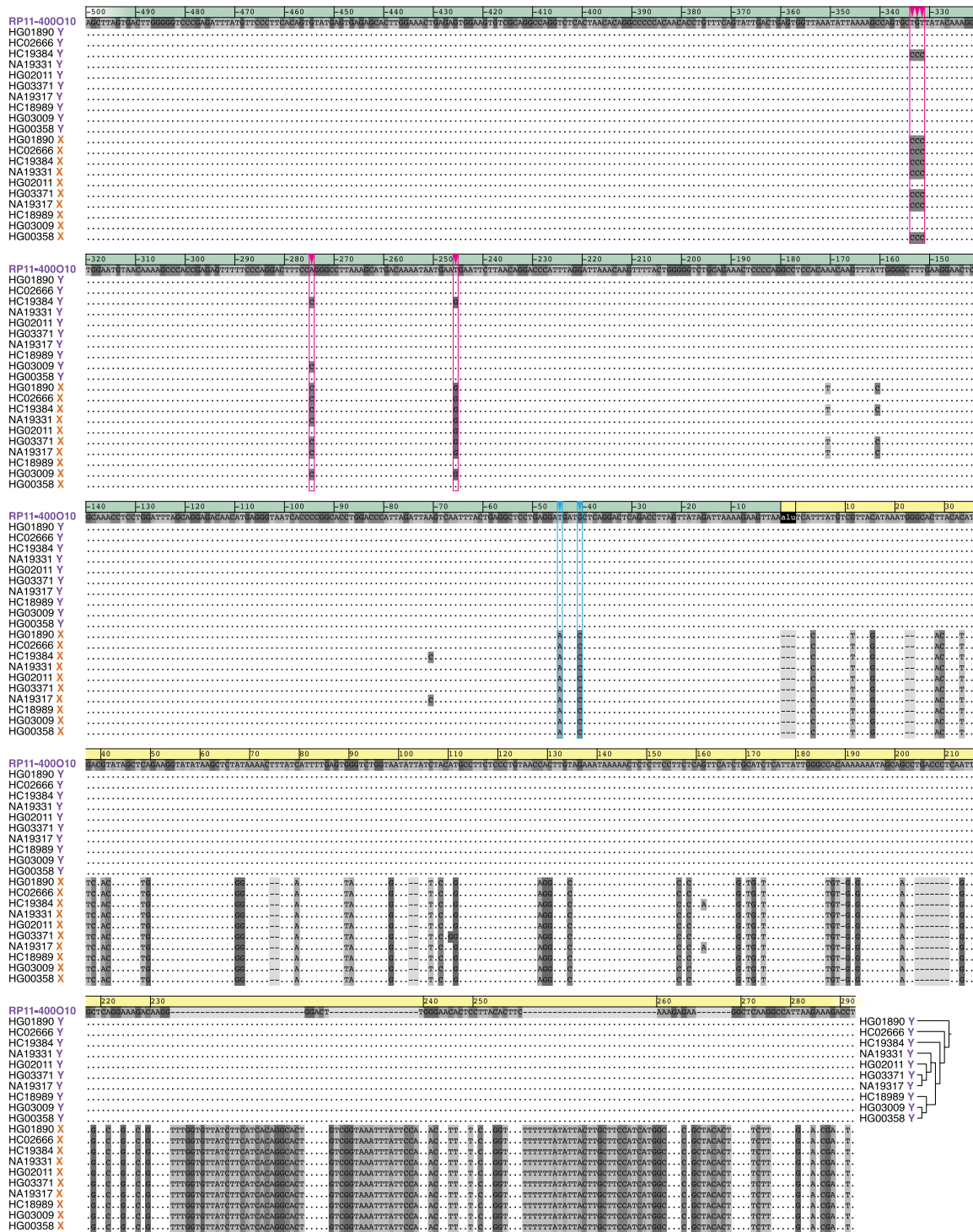


Figure 7. Polymorphisms shared by X and Y chromosomes within 500 bases of the 1989 PAB

Multiple alignment of human sex chromosome sequences showing the 1989 PAB at or near the Y-specific Alu insertion (black). Dots indicate identity with RP11-400O10 clone from human Y; coordinates relative to Alu insertion. Tree indicates NPY haplogroup tree for ten contiguous Y chromosome sequences produced by Hallast and colleagues.⁸ Magenta arrowheads and boxes mark polymorphisms shared by X and Y; cyan arrowheads and boxes mark fixed differences.

Initiative, and generous gifts from The Brit Jepson d'Arbeloff Center on Women's Health, Arthur W. and Carol Tobin Brill, Matthew Brill, Charles Ellis, the Baltimore Bar Foundation, the Howard P. Colhoun Family Foundation, the Simons Foundation, and the Knobloch Family Foundation.

Author contributions

Conceptualization: D.W.B., J.H., H.S., and D.C.P.; data curation, formal analysis, methodology, software, and validation: D.W.B., J.H., H.S., and E.C.O.; visualization: D.W.B. and E.C.O.; funding

acquisition, project administration, supervision: D.C.P.; writing – original draft: D.W.B.; writing – review & editing: D.W.B., J.H., H.S., E.C.O., and D.C.P.

Declaration of interests

The authors declare no competing interests.

Web resources

1000 Genomes Project, phased individual-level PAR variants, http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20201028_3202_phased_CCDG_14151_B01_GRM_WGS_2020-08-05_chrX.filtered.eagle2-phased.v2.vcf.gz

Database of Genotypes and Phenotypes (dbGaP), Single-Sperm Genome Sequencing of Sperm Donors, study phs001887, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001887.v1.p1

GenBank, <https://www.ncbi.nlm.nih.gov/genbank/>

NCBI Sequence Read Archive, single-sperm genome sequencing of sperm donors, study phs001887, https://www.ncbi.nlm.nih.gov/sra?linkname=bioproject_sra_all&from_uid=561434

RepeatMasker, <https://www.repeatmasker.org/>

Rutgers Combined Linkage-Physical Map, http://compgen.rutgers.edu/rutgers_maps.shtml

Y chromosome assemblies from Hallast and colleagues, http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HGSVC3/working/20230412_sigY_assembly_data

Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2024.09.005>.

Received: June 8, 2024

Accepted: September 24, 2024

Published: October 14, 2024

References

1. Darlington, C.D., Haldane, J.B.S., and Roller, P.C. (1934). Possibility of Incomplete Sex Linkage in Mammals. *Nature* *133*, 417. <https://doi.org/10.1038/133417b0>.
2. Koller, P.C., and Darlington, C.D. (1934). The genetical and mechanical properties of the sex-chromosomes. *J. Genet.* *29*, 159–173. <https://doi.org/10.1007/bf02982193>.
3. Burgoyne, P.S. (1982). Genetic homology and crossing over in the X and Y chromosomes of mammals. *Hum. Genet.* *61*, 85–90. <https://doi.org/10.1007/BF00274192>.
4. Goodfellow, P. (1983). Expression of the 12E7 Antigen is Controlled Independently by Genes on the Human X and Y Chromosomes. In *Mechanisms of Gonadal Differentiation in Vertebrates* (Springer Berlin Heidelberg), pp. 35–39.
5. Goodfellow, P.J., Darling, S.M., Thomas, N.S., and Goodfellow, P.N. (1986). A Pseudoautosomal Gene in Man. *Science* *234*, 740–743.
6. Kvaløy, K., Galvagni, F., and Brown, W.R. (1994). The sequence organization of the long arm pseudoautosomal region of the human sex chromosomes. *Hum. Mol. Genet.* *3*, 771–778. <https://doi.org/10.1093/hmg/3.5.771>.
7. Ellis, N.A., Goodfellow, P.J., Pym, B., Smith, M., Palmer, M., Fritschauf, A.-M., and Goodfellow, P.N. (1989). The pseudoautosomal boundary in man is defined by an Alu repeat sequence inserted on the Y chromosome. *Nature* *337*, 81–84. <https://doi.org/10.1038/337081a0>.
8. Hallast, P., Ebert, P., Loftus, M., Yilmaz, F., Audano, P.A., Logsdon, G.A., Bonder, M.J., Zhou, W., Höps, W., Kim, K., et al. (2023). Assembly of 43 human Y chromosomes reveals extensive complexity and variation. *Nature* *621*, 355–364. <https://doi.org/10.1038/s41586-023-06425-6>.
9. Cooke, H.J., Brown, W.R., and Rappold, G.A. (1985). Hyper-variable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature* *317*, 687–692. <https://doi.org/10.1038/317687a0>.
10. Simmler, M.-C., Rouyer, F., Vergnaud, G., Nyström-Lahti, M., Ngo, K.Y., de la Chapelle, A., and Weissenbach, J. (1985). Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature* *317*, 692–697. <https://doi.org/10.1038/317692a0>.
11. Rouyer, F., Simmler, M.-C., Johnsson, C., Vergnaud, G., Cooke, H.J., and Weissenbach, J. (1986). A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* *319*, 291–295. <https://doi.org/10.1038/319291a0>.
12. Page, D.C., Bieker, K., Brown, L.G., Hinton, S., Leppert, M., Lalouel, J.-M., Lathrop, M., Nyström-Lahti, M., de la Chapelle, A., and White, R. (1987). Linkage, physical mapping, and DNA sequence analysis of pseudoautosomal loci on the human X and Y chromosomes. *Genomics* *1*, 243–256. [https://doi.org/10.1016/0888-7543\(87\)90051-6](https://doi.org/10.1016/0888-7543(87)90051-6).
13. Henke, A., Fischer, C., and Rappold, G.A. (1993). Genetic map of the human pseudoautosomal region reveals a high rate of recombination in female meiosis at the Xp telomere. *Genomics* *18*, 478–485. [https://doi.org/10.1016/s0888-7543\(11\)80003-0](https://doi.org/10.1016/s0888-7543(11)80003-0).
14. Pritchard, C.A., Goodfellow, P.J., and Goodfellow, P.N. (1987). Mapping the limits of the human pseudoautosomal region and a candidate sequence for the male-determining gene. *Nature* *328*, 273–275. <https://doi.org/10.1038/328273a0>.
15. Ellis, N., Yen, P., Neiswanger, K., Shapiro, L.J., and Goodfellow, P.N. (1990). Evolution of the pseudoautosomal boundary in old world monkeys and great apes. *Cell* *63*, 977–986. [https://doi.org/10.1016/0092-8674\(90\)90501-5](https://doi.org/10.1016/0092-8674(90)90501-5).
16. Ellis, N., Taylor, A., Bengtsson, B.O., Kidd, J., Rogers, J., and Goodfellow, P. (1990). Population structure of the human pseudoautosomal boundary. *Nature* *344*, 663–665. <https://doi.org/10.1038/344663a0>.
17. Matisse, T.C., Chen, F., Chen, W., De La Vega, F.M., Hansen, M., He, C., Hyland, F.C.L., Kennedy, G.C., Kong, X., Murray, S.S., et al. (2007). A second-generation combined linkage-physical map of the human genome. *Genome Res.* *17*, 1783–1786. <https://doi.org/10.1101/gr.7156307>. Epub 2007 Nov 7.
18. Flaquer, A., Fischer, C., and Wienker, T.F. (2009). A New Sex-Specific Genetic Map of the Human Pseudoautosomal Regions (PAR1 and PAR2). *Hum. Hered.* *68*, 192–200. <https://doi.org/10.1159/000224639>.
19. Hinch, A.G., Altemose, N., Noor, N., Donnelly, P., and Myers, S.R. (2014). Recombination in the Human Pseudoautosomal Region PAR1. *PLoS Genet.* *10*, e1004503. <https://doi.org/10.1371/journal.pgen.1004503>.
20. Schmitt, K., Lazzeroni, L.C., Foote, S., Vollrath, D., Fisher, E.M., Goradia, T.M., Lange, K., Page, D.C., and Arnheim, N. (1994). Multipoint linkage map of the human

- pseudoautosomal region, based on single-sperm typing: do double crossovers occur during male meiosis? *Am. J. Hum. Genet.* *55*, 423–430.
21. Lien, S., Szyda, J., Schechinger, B., Rappold, G., and Arnheim, N. (2000). Evidence for Heterogeneity in Recombination in the Human Pseudoautosomal Region: High Resolution Analysis by Sperm Typing and Radiation-Hybrid Mapping. *Am. J. Hum. Genet.* *66*, 557–566. <https://doi.org/10.1086/302754>.
 22. Lu, S., Zong, C., Fan, W., Yang, M., Li, J., Chapman, A.R., Zhu, P., Hu, X., Xu, L., Yan, L., et al. (2012). Probing Meiotic Recombination and Aneuploidy of Single Sperm Cells by Whole-Genome Sequencing. *Science* *338*, 1627–1630.
 23. Löytynoja, A., and Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. U.S.A.* *102*, pp. 10557–10562. <https://doi.org/10.1073/pnas.0409137102>.
 24. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* *59*, 307–321. <https://doi.org/10.1093/sysbio/syq010>.
 25. Byrsk-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., Corvelo, A., Clarke, W.E., Musunuri, R., Nagulapalli, K., et al. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* *185*, 3426–3440.e19. <https://doi.org/10.1016/j.cell.2022.08.004>.
 26. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359. <https://doi.org/10.1038/nmeth.1923>.
 27. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* *25*, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
 28. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* *31*, 2032–2034. <https://doi.org/10.1093/bioinformatics/btv098>.
 29. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1207.3907>.
 30. Garrison, E., Kronenberg, Z.N., Dawson, E.T., Pedersen, B.S., and Prins, P. (2022). A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PLoS Comput. Biol.* *18*, e1009123. <https://doi.org/10.1371/journal.pcbi.1009123>.
 31. Bell, A.D., Mello, C.J., Nemes, J., Brumbaugh, S.A., Wysoker, A., and McCarroll, S.A. (2020). Insights into Variation in Meiosis from 31,228 Human Sperm Genomes. *Nature* *583*, 259–264. <https://doi.org/10.1038/s41586-020-2347-0>.
 32. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* *27*, 573–580. <https://doi.org/10.1093/nar/27.2.573>.
 33. Darling, A.E., Mau, B., and Perna, N.T. (2010). progressive-Mauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* *5*, e11147. <https://doi.org/10.1371/journal.pone.0011147>.
 34. Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: Maximum Likelihood Phylogenetic Analysis Using Quartets and Parallel Computing. *Bioinformatics* *18*, 502–504. <https://doi.org/10.1093/bioinformatics/18.3.502>.
 35. R Core Team (2020). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing).
 36. Okonechnikov, K., Golosova, O., Fursov, M.; and UGENE team (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* *28*, 1166–1167. <https://doi.org/10.1093/bioinformatics/bts091>.
 37. Shen, M.R., Batzer, M.A., and Deininger, P.L. (1991). Evolution of the master Alu gene(s). *J. Mol. Evol.* *33*, 311–320. <https://doi.org/10.1007/BF02102862>.
 38. Hook, E.B., and Warburton, D. (2014). Turner syndrome revisited: Review of new data supports the hypothesis that all viable 45,X cases are cryptic mosaics with a rescue cell line, implying an origin by mitotic loss. *Hum. Genet.* *133*, 417–424. <https://doi.org/10.1007/s00439-014-1420-x>.
 39. Lahn, B.T.; and Page (1999). Four Evolutionary Strata on the Human X Chromosome. *Science* *286*, 964–967.
 40. Bellott, D.W., Hughes, J.F., Skaletsky, H., Brown, L.G., Pyntikova, T., Cho, T.-J., Koutseva, N., Zaghlul, S., Graves, T., Rock, S., et al. (2014). Mammalian Y Chromosomes Retain Widely Expressed Dosage-Sensitive Regulators. *Nature* *508*, 494–499. <https://doi.org/10.1038/nature13206>.
 41. Kumar, S., Stecher, G., Suleski, M., and Heddes, S.B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* *34*, 1812–1819. <https://doi.org/10.1093/molbev/msx116>.