Where is the boundary of the human pseudoautosomal region?

Daniel W. Bellott[1], Jennifer Hughes[1], Helen Skaletsky[1,2],

Erik C. Owen[1,3], and David C. Page[1,2,4] *

1. Whitehead Institute, Cambridge, Massachusetts 02142, USA;
2. Howard Hughes Medical Institute, Whitehead Institute, Cambridge, Massachusetts 02142, USA;
3. Program in Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA, 02142, USA
4. Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
* Correspondence: dcpage@wi.mit.edu

1

Abstract

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[1]. 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 the sequence assemblies presented by Hallast and colleagues 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[2,3]. 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 percent[3]. 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[4]. 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[4]. 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[5,6].

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

57 chromosome[7] (Figure 1a). The pseudoautosomal boundary (PAB) is the point on Xp and Yp
58 where crossing-over between them ceases and the allelic PARs give way to divergent NPX
59 and NPY sequences. The localization of *CD99* provided a foothold for further
60 characterization of PAR1, and, in 1989, Ellis and colleagues fine mapped the PAB of PAR1,
61 identifying an Alu repeat sequence inserted on the Y chromosome as the PAB[8] (Figure
62 1b&c). This identification was supported by complementary genetic and physical mapping
63 techniques. Pedigree-based genetic linkage maps, linkage-disequilibrium-based maps, and
64 single sperm typing revealed patterns of crossing-over between X and Y chromosomes,
65 while somatic cell hybrids, radiation-hybrid mapping, Southern blots, and clone-based
66 sequencing all established physical linkage between the PAR1 and X- or Y-specific DNA
67 sequences.

68 Hallast and colleagues recently employed a novel approach to locate the PAB of PAR1, by
69 using sequence variation in the human population as a proxy for crossing-over[1]. The
70 recombination rate varies across the human genome, and regions of high recombination
71 are observed to have a higher density of structural variants and indels. Both recombination
72 and variation are highest in PAR1, followed by the autosomes, the NPX, and are lowest in
73 the NPY. Hallast and colleagues observed that the 500-kb segment immediately distal to the
74 1989 PAB shows a low density of structural variants and indels among their sequenced Y
75 chromosomes, and they reasoned that this observation is best explained by reduced or
76 absent recombination in this interval. Hallast and colleagues concluded:

77   "our data suggest that the boundary between the recombining pseudoautosomal
78   region 1 and the non-recombining portions of the X and Y chromosomes lies
79   500 kb away from the currently established boundary"

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

Establishing the PAB

87 The earliest genetic maps of PAR1 were based on analysis of human pedigrees, using
88 cloned DNA sequences as probes to identify restriction fragment length polymorphisms on
89 Southern blots[6,9–13]. Rouyer and colleagues genotyped three pseudoautosomal markers
90 across eight Centre d'Etude du Polymorphism Humain (CEPH) families to produce the first
91 genetic linkage map of PAR1[11]. The map revealed a gradient of sex linkage from the most
92 proximal to the most telomeric marker, and covered a distance of 50 cM (centiMorgans),
93 indicating that, on average, one crossover occurs in PAR1 during each male meiosis[11].
94 However, *DXYS17*, the marker most closely linked to sex, was still 14 cM distal to the PAB
95 (Figure 1b). Subsequently, Goodfellow and colleagues re-mapped these markers, along
96 with the *CD99* gene, across an independent set of three pedigrees[6]. *CD99* was the proximal
97 marker, mapping only 2.2 cM distal to biological sex[6]. Thus, the PAB for PAR1 must lie
98 between *CD99* and the sex determining locus.

99   The tight linkage of *CD99* with sex inspired researchers to attempt to isolate the boundary
100   between PAR1 and the NPY using *CD99* as a probe. Using a 49,XYYYY cell line as a highly
101   enriched source of Y-linked DNA, Pritchard and colleagues used rare-cutting restriction
102   enzymes, both singly and in pairs, to generate long restriction fragments which they
103   resolved by pulsed-field gel electrophoresis[14]. They identified a 200-kb SfiI/ClaI fragment
104   that hybridized to both *CD99* and NPY sequences[14]. Subsequently, Ellis and colleagues
105   further refined this interval, beginning with a 110-kb restriction fragment containing their
106   most proximal *CD99* and most distal NPY probes, and gradually narrowed this down to a
107   0.8-kb restriction fragment containing the boundary between PAR1 and NPY[8]. Armed with
108   precise knowledge of the most proximal PAR sequences, they also isolated a 0.9-kb
109   restriction fragment containing the boundary between PAR1 and NPX[8]. They sequenced
110   both fragments and found they were 99% identical distal to a 303-bp Alu repeat insertion
111   in the Y-derived sequence[8]. Proximal to the Alu insertion, X-Y identity drops to 77% for
112   another 245 bases, and fails to align further[8]. They reasoned that the high X-Y identity in
113   the sequence distal to the Alu insertion was maintained by crossing-over, and that the
114   distal end of the Alu insertion marked the PAB (Figure 1c).

115   To establish that this sequence did constitute the PAB, it was critical to find additional
116   evidence of crossing-over. Ellis and colleagues took a two-pronged approach, examining
117   homologous X and Y sequences across catarrhine primates and sequencing this region from
118   diverse human populations to identify polymorphisms[15,16]. Their evolutionary analysis
119   found that while the Alu insertion was found across the Y chromosomes of great apes, it
120   was absent from Old World monkeys[15]. However, even Old World monkeys maintained a
121   crisp drop-off in X-Y identity between regions distal and proximal to the Alu insertion
122   site[15]. This suggested that X-Y identity in the distal region was maintained by crossing-over
123   since the split of Old World monkeys and great apes. By surveying 57 Y chromosomes and
124   60 X chromosomes from ten human populations, Ellis and colleagues identified single-
125   nucleotide polymorphisms to infer patterns of crossing-over[16]. While there were fixed X-Y
126   differences at positions 41 and 45 bases distal to the Alu insertion, they observed a site
127   with a polymorphism shared by the X and the Y only 274 bases distal to the Alu[16]. Between
128   these two sites, all Y chromosomes shared SNPs with a common X chromosome haplotype.
129   This provided strong evidence that crossovers had occurred between 45 and 274 bases
130   distal to the Alu insertion, localizing the human PAB to this 229-base interval.

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

138   In producing the second iteration of the Rutgers Map of the human genome, Matise and
139   colleagues mapped 28,121 markers on CEPH pedigrees, including three markers within
140   *CD99* (rs924904, rs310136, and rs311071)[17] (Figure 1c). These markers mapped from 5.12
141   to 2.79 cM distal to the 1989 PAB[17]. Flaquer and colleagues added more pseudoautosomal
142   markers to the third iteration of the Rutgers map, and mapped two of these markers even

143   closer to the PAB: rs310136, 87.6 kb and 0.7 cM away; and rs311071, 66.3 kb and 0.0007
144   cM distal to the PAB[18,19]. Subsequently, Hinch and colleagues generated a fine-scale map of
145   the PAR; they observed 126 paternal crossovers in PAR1, of which at least 14 were in the
146   500 kb immediately distal to the 1989 PAB[20]. This region covers 5.1 cM in their map, and
147   includes a recombination hotspot within the *XG* gene, which straddles the PAB on the X
148   chromosome, but is disrupted on the Y chromosome[20].

149   Schmitt and colleagues typed 903 flow-sorted sperm from two donors to map
150   recombination among five loci in PAR1, and used 2,005 sperm from a third donor to map
151   *DXYS77*, the marker closest to the 1989 PAB, at higher resolution[21]. *DXYS77* is located in an
152   intron of the *XG* gene, 19.6 kb and 0.77 cM distal to the PAB (Figure 1c). Subsequent single-
153   sperm typing analyses have not mapped PAR1 at such high resolution, but have also
154   observed crossovers in the 500 kb distal to the 1989 PAB[22,23].

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

168   ## Methods

169   ### Primate X-Y alignments

170   We downloaded primate X- and Y-linked sequences spanning the PAB from GenBank
171   (Supplemental Table 1). Prior to alignment, we removed the Alu insertion at the 1989 PAB
172   from ape Y-linked sequences, as well as an independent Alu insertion 300 bp proximal to
173   the 1989 PAB in the colobus NPY. We used PRANK (v.170427)[24] to generate a multiple
174   alignment of sequences distal and proximal to the 1989 PAB (Supplemental Data 1). We
175   used various windows of these multiple alignments to infer the phylogenetic relationships
176   between sequences with PhyML (v.3.3.20190909)[25] with 100 bootstrap replicates.

177   ### Human polymorphisms from 1000 genomes

178   We downloaded phased individual-level variants from the PAR from the 1000 Genomes
179   Project[26]. For each of 118,327 polymorphic sites in PAR1, we tabulated the number of
180   males and females who were homozygous for the reference allele, heterozygous, and
181   homozygous for the alternative allele (Supplemental Table 2). We used these counts to
182   calculate sex-specific minor allele frequencies for males and females.

183 **Single sperm sequencing**

184 We obtained raw sequence data dbGaP study phs001887 from the NCBI Sequence Read
185 Archive. Using Bowtie2 (version 2.4.2)[27], we aligned reads to a version of hg38
186 (GCF_000001405.40) with the chrY PAR masked. For each donor, we combined PAR
187 alignments from all haploid cells using samtools (version 1.11)[28], marked duplicate reads
188 with sambamba (version 0.8.2)[29], and called variants at sites in dbSNP with freebayes
189 (version 1.3.2)[30]. We used vcffilter[31] to restrict downstream analyses to biallelic SNPs with
190 Q > 20, more than 2 reads each of reference and alternate alleles, and allele frequencies
191 between 0.25 and 0.75. We identified X- and Y-bearing sperm by the depth of reads
192 mapping to NPX and NPY sequences. In our analyses of the Sperm-seq data from
193 phs001887, we also excluded reads with cell barcodes associated with cell doublets and
194 bead doublets by Bell and colleagues[32].

195 For each donor, we reconstructed X and Y haplotypes across PAR1 in two passes. In each
196 pass we assigned the alleles at each variant site to either the X or Y donor haplotype to
197 minimize the number of crossovers with more proximal sites, then, within each sample,
198 localized crossovers between sites with variants assigned to different donor haplotypes. In
199 the first pass, we flagged markers in tight double recombinants (closer than 278148 bases;
200 that is, one tenth the length of PAR1, or ~5 cM), as unreliable, and identified non-crossover
201 samples. In the second pass, we excluded unreliable markers flagged in the first pass, as
202 well as any markers with more than 10% discordant genotypes in non-crossover samples.
203 We then removed any remaining tight double recombinants before calculating the local
204 recombination rate.

205 **Reanalysis of Y chromosomes sequenced by Hallast and colleagues**

206 Hallast and colleagues produced contiguous assemblies of the PAR-NPY boundary from ten
207 samples (HG01890, HC02666, HC19384, NA19331, HG02011, HG03371, NA19317,
208 HC18989, HG03009, and HG00358); we downloaded their assemblies
209 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HGSVC3/working/20230412_
210 sigY_assembly_data), and extracted the PAR contigs for further analysis.

211 To analyze clock-like evolution on the Y chromosome, we first masked simple repeats with
212 RepeatMasker (version 4.1.5)[33] and tandem repeats with trf (version 4.09)[34], then aligned
213 all ten Y chromosome assemblies with progressiveMauve (build date May 15 2023 at
214 08:46:41)[35]. We extracted 20-kb windows of the multiple alignment, excluding columns of
215 gaps and masked characters, at 1-kb intervals. We used tree-puzzle (version 5.2)[36] to
216 conduct a likelihood ratio test to compare a model where sequences evolve using a single
217 clock-like rate based on their relationships from the Y chromosome haplogroup tree,
218 versus a model where rates are allowed to vary across branches. We obtained p-values
219 from the test statistic reported by tree-puzzle (Supplemental Table 4), using the `pchisq`
220 function and adjusted for multiple testing by Bonferroni correction with the `p.adjust`
221 function in R (version 4.1.1)[37].

222 To look for polymorphisms shared among X and Y chromosomes, we aligned homologous X
223 and Y sequences from the ten genomes with contiguous PAR sequence and RP11-400O10, a

224 human Y chromosome BAC spanning the 1989 PAB, and manually edited the resulting
225 alignments in Unipro UGENE (version 49.1)[38] (Supplemental Data 3)).

Results

Primate X-Y alignments

228 We took advantage of the rich genomic resources from multiple species to revisit the
229 evolutionary analyses of Ellis and colleagues, using high-quality assemblies of primate X
230 and Y chromosomes. Within the PAR, crossing-over will result in the DNA sequences of Y
231 chromosomes being more similar to X chromosomes of the same species than to Y
232 chromosomes of other species. Conversely, in the NPY, the absence of crossing-over will
233 result in NPY sequences from different species being more similar to each other than to
234 NPX sequences from the same species. We curated high-quality sequences adjacent to the
235 1989 PAB from the NPX and NPY from seven catarrhine primate species: human,
236 chimpanzee, gorilla, orangutan, siamang, rhesus macaque, and colobus monkey
237 (Supplemental Table 1, Supplemental Data 1). We constructed a multiple alignment from
238 these sequences, and used these alignments to construct trees using a range of window
239 sizes from the alignment, both distal and proximal to the 1989 PAB (Figure 2). Across all
240 window sizes, even as narrow as 1 kb from the 1989 PAB (Figure 2d), we observe that
241 sequences distal to the 1989 PAB group together by species, while sequences proximal to
242 the 1989 PAB group together by chromosome. Examining this multiple alignment more
243 closely, we find human X and Y chromosomes sharing an A>G substitution only 93 bases
244 distal to the 1989 PAB (Figure 3). Since this substitution is also present on the chimpanzee
245 Y, this suggests that it originated on the Y chromosome in the last common ancestor of
246 humans and chimpanzees, and crossed over to the X chromosome in humans (Figure 3).
247 The Alu element, from the cattarhine-specific AluY family[39], was inserted into the Y
248 chromosome after apes diverged from Old World monkeys, and has not crossed over to the
249 X. This Alu insertion must have occurred at or near the ancestral catarrhine PAB; only 11
250 bases proximal to this site, all X chromosomes share a T where all Y chromosomes carry a G
251 or A (Figure 3), confirming that crossing-over was suppressed in this proximal region
252 before the divergence of apes and Old World monkeys.

Human polymorphisms from 1000 Genomes

254 To gain additional insight into the history of crossing-over within the human population,
255 we analyzed polymorphism data from the 1000 Genomes Project[26]. Strictly Y-linked
256 regions with low or absent crossing-over with the X chromosome will accumulate
257 mutations that are male-specific, leading to sex biases in the frequency of minor alleles.
258 Using data from 1,599 males and 1,603 females, we calculated sex-specific minor allele
259 frequencies for 118,327 polymorphic sites across PAR1 (Supplemental Table 2). We
260 plotted the sex difference in minor allele frequency throughout PAR1 and observed that the
261 differences between male and female minor allele frequencies are typically small, even in
262 the 500 kb immediately distal to the 1989 PAB (Figure 4a). Within 10 kb of the 1989 PAB,
263 alleles are more sex-biased, and this sex bias in minor allele frequencies increases rapidly
264 as one moves proximally, nearing the maximum possible value of 0.5 (Figure 4a). Distal to
265 this 10-kb segment adjacent to the 1989 PAB, male and female minor allele frequencies

266 closely mirror each other, indicating that crossing-over has intermixed and homogenized
267 the sets of alleles on the X and Y chromosomes (Figure 4b).

268 ## Single sperm sequencing

269 To capture ongoing crossover events in PAR1 from living individuals, we analyzed a
270 massive single-sperm sequencing dataset (dbGaP project phs001887), containing Illumina
271 reads from 31,228 sperm across 20 donors, which has not been previously analyzed for
272 PAR1 crossovers[32]. We identified X- and Y- bearing sperm from the depth of reads aligning
273 to the NPX and NPY and used this information to phase variant sites in PAR 1. Across all
274 donors, we observed 795 sperm with unambiguous crossovers between the 1989 and 2023
275 PABs (Supplemental Table 3). We calculated the recombination rate between all genotyped
276 variants in PAR1 (Figure 5a, Supplemental Figure 1), and estimate the total genetic
277 distance between the 1989 PAB and 2023 PAB to be 4.78 cM (Figure 5b). We observe a
278 local increase in the recombination rate close to the 1989 PAB (Figure 5a), parallel what
279 Hinch and colleagues observed using a pedigree-based genetic map[20]. We note that this
280 local pattern of increased recombination near the 1989 PAB is not present across all
281 donors (Supplemental Figure 1), consistent with the observation that PAR1 hotspots are
282 rapidly evolving and differ among human populations[20].

283 ## Reanalysis of Y chromosomes sequenced by Hallast and colleagues

284 All three of these fully independent analyses confirmed the location of the 1989 PAB, so we
285 asked whether the sequences generated by Hallast and colleagues[1] might also provide
286 evidence for crossing-over in the 500 kb distal to the 1989 PAB. Of the 43 individual human
287 male genomes assembled by Hallast and colleagues, ten have contiguous PAR1 assemblies,
288 supported by high-quality differences between long reads originating from the X and Y
289 chromosomes[1]. We restricted our analyses to these ten genomes where the provenance of
290 the sequences within the PAR are unambiguous.

291 During human evolution, substitutions are predicted to accumulate on the NPY in a clock-
292 like fashion, with no recombination. However, substitutions that arise in the PAR can be
293 transferred from Y chromosomes to X chromosomes, or from one Y chromosome
294 haplogroup to another through an X-chromosome intermediate, across generations. As a
295 result, for sequences within the PAR, branch lengths calculated using the NPY haplogroup
296 tree will show deviations from clock-like behavior. We constructed a multiple alignment of
297 all ten contiguous Y-linked PARs (Supplemental Data 2). Using the phylogenetic
298 relationships between these sequences based on their NPY haplogroup[1], we conducted a
299 likelihood ratio test (LRT) in 20-kb sliding windows of the alignment, comparing a single-
300 clock model where all branches accumulate substitutions at the same rate, versus a local-
301 clock alternative where rates are allowed to vary among branches (Figure 6, Supplemental
302 Table 4). We observe a crisp change in the results of this test that coincides with the 1989
303 PAB (Figure 6). Windows proximal to the 1989 PAB are almost universally consistent with
304 clock-like evolution along the haplogroup tree, as expected for NPY sequences. The rare
305 exceptions may represent X-Y gene conversion events in homologous sequences. Distal to
306 the 1989 PAB, the single-clock model can be rejected in the majority of windows. This can

8

307   readily be explained by a history of crossing-over distal to the 1989 PAB, so that the Y-
308   haplogroup tree does not accurately describe the relationship between these sequences.

309   Encouraged by these results, we further examined the homologous X and Y sequences from
310   these ten genomes to ask how close to the 1989 PAB we could observe evidence of
311   crossing-over. Following the procedure of Ellis and colleagues, we constructed a multiple
312   alignment of these ten diverse X- and Y-linked sequences, along with sequence from a BAC
313   clone from the Y chromosome spanning the 1989 PAB (Figure 7, Supplemental Data 3), and
314   looked for polymorphisms shared by X and Y chromosomes. We observe the same fixed
315   differences between human X and Y chromosomes at positions 41 and 45 bases distal to
316   the Alu insertion that Ellis and colleagues described more than 30 years ago (Figure 7).
317   Similarly, we see that all Y chromosomes share a haplotype commonly found on X
318   chromosomes up to 245 bases distal to the Alu insertion (Figure 7). We observe five
319   previously reported polymorphic sites (rs147962110, rs189310269, rs180883975,
320   rs2534635, and rs5982851) where reference and alternate alleles are shared by X and Y
321   chromosomes (Figure 7). The closest site, rs5982851, is only 246 bases distal to the Alu
322   insertion that marks the PAB (Figure 7), while the next closest, rs2534635, was previously
323   observed by Ellis and colleagues[16]. Thus, rather than overturning the 1989 PAB, the
324   sequences produced by Hallast and colleagues reveal a historical recombination event 28
325   bases closer to the 1989 PAB than Ellis and colleagues initially observed, narrowing the
326   possible location of the PAB to the 201-base interval between rs5982851 and the first fixed
327   difference between the X and Y chromosomes.

Discussion

329   Our present understanding of the biology of the sex chromosomes has been informed by
330   close study of crossing-over between the X and Y chromosomes. During meiosis, crossing-
331   over generates a physical link between each pair of homologous chromosomes, providing
332   the tension required to ensure that homologs can be segregated faithfully. The PARs are
333   responsible for providing this obligate crossover between the X and Y chromosomes during
334   male meiosis. Without this crossover, meiosis generates XY sperm that produce 47,XXY
335   offspring that develop Klinefelter syndrome. In contrast, when there are crossovers
336   involving the NPX or NPY, sperm will carry structurally anomalous X and Y chromosomes
337   that produce 46,XX male (DSD) and 46,XY female (DSD) offspring. All three conditions
338   result in infertility, imposing strong selection to maintain the PAR. The highly
339   differentiated human X and Y chromosomes evolved from what were once an ordinary pair
340   of autosomes that crossed over freely. This divergence was the result of a series of
341   rearrangements over the last 200 million years that first suppressed crossing-over around
342   *SRY*, the testis-determining gene, initially creating a PAB, and then repeatedly shifted the
343   PAB to expand the NPY at the expense of the PAR[40]. The absence of crossing-over in males
344   radically altered both the NPX and NPY; 97% of ancestral NPY genes were lost to genetic
345   decay, while their counterparts on the NPX evolved new regulatory mechanisms to
346   maintain the ancestral gene dose in both sexes[41]. As a result, the PAB forms the frontier
347   between the sex-specific biology of the NPX and NPY, and the sex-shared biology of the
348   PAR. As we continue to investigate the genetics of sex differences, the PAB remains a
349   critical landmark on the human sex chromosomes.

350 For 40 years the PAB has been defined by the pattern of crossing-over. The available
351 evidence overwhelmingly supports the 1989 PAB in PAR1, rather than a site 500 kb more
352 distal. Crossing-over between the X and Y chromosomes has maintained high X-Y identity
353 across the region distal to the 1989 PAB since the divergence of apes and Old World
354 monkeys, around 30 million years ago[42]. Within the human population, crossovers
355 occurred as close as 246 bases to the 1989 boundary, as inferred from alleles at
356 polymorphic sites shared by both X and Y chromosomes. Human pedigrees and single-
357 sperm typing data record ongoing crossovers throughout the PAR, including the 500 kb
358 adjacent to the 1989 boundary. It is therefore clear that that the absence of structural
359 variation cannot be used as a proxy for the absence of crossing-over. However, we note that
360 our findings do not otherwise challenge the methods or findings of Hallast and colleagues,
361 which in other respects agree closely with previous characterizations of the human Y
362 chromosome.

Declaration of interests

364 The authors declare no competing interests.

Acknowledgments

Author Contributions

372 Conceptualization: D.W.B., J.H., H.S., D.C.P.; Data curation, Formal Analysis, Methodology,
373 Software, Validation: D.W.B., J.H., H.S., E.C.O.; Visualization: D.W.B., E.C.O.; Funding
374 acquisition, Project administration, Supervision: D.C.P.; Writing – original draft: D.W.B.;
375 Writing – review & editing: D.W.B., J.H., H.S., E.C.O., D.C.P.

Data and Code Availability

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

379   ## Figures and Legends

380   **Figure 1: The boundary of human PAR1.**

381   Schematic representation of the human X and Y chromosomes, (a) showing the whole of
382   each chromosome, including both PARs. Shading indicates four classes of euchromatic
383   sequence, as well as heterochromatic sequences. Colored arcs indicate regions of >99% X-Y
384   sequence identity. (b) Distal 4 Mb of the short arm of each chromosome, enlarged to show
385   PAR1. Dashed lines: 2023 and 1989 PAB. Boxes: genes in PAR1, NPX, and NPY.
386   Arrowheads: genetic markers. (c) 250 kb of each chromosome near the 1989 PAB. Genes
387   shown as lines with arrowheads indicating direction of transcription, exons as boxes. *XG*
388   and *XGY2* share three exons in the PAR.

389

11
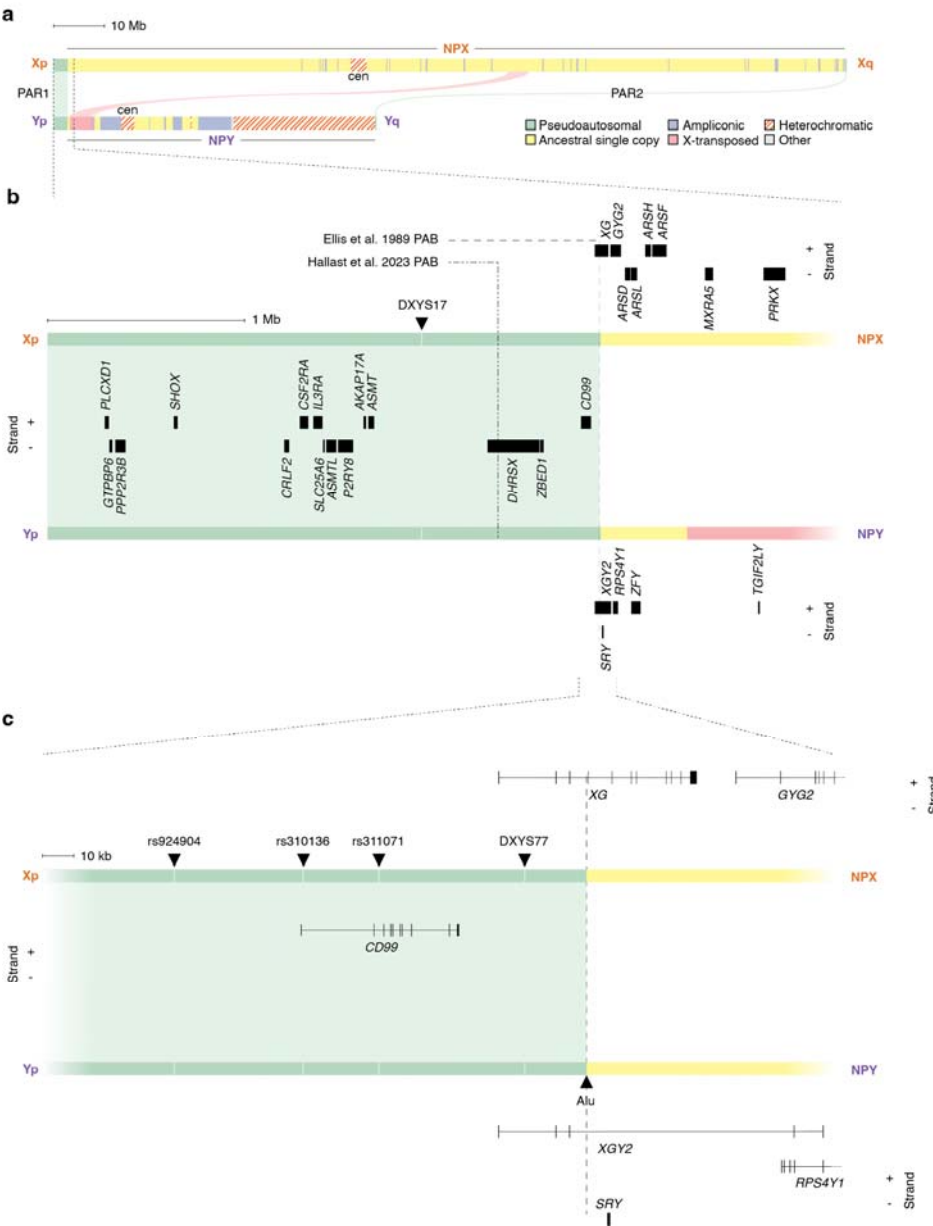
390   **Figure 2: Phylogenetic relationships between primate X and Y sequences differ distal and**
391   **proximal to 1989 PAB.**

392   Maximum likelihood trees based on multiple alignment of primate sex chromosome
393   sequences in windows (a) 20 kb, (b) 10 kb, (c) 5 kb, or (d) 1 kb upstream and downstream
394   of the 1989 PAB; nodes show support for branching order from 100 bootstrap replicates.

395

396
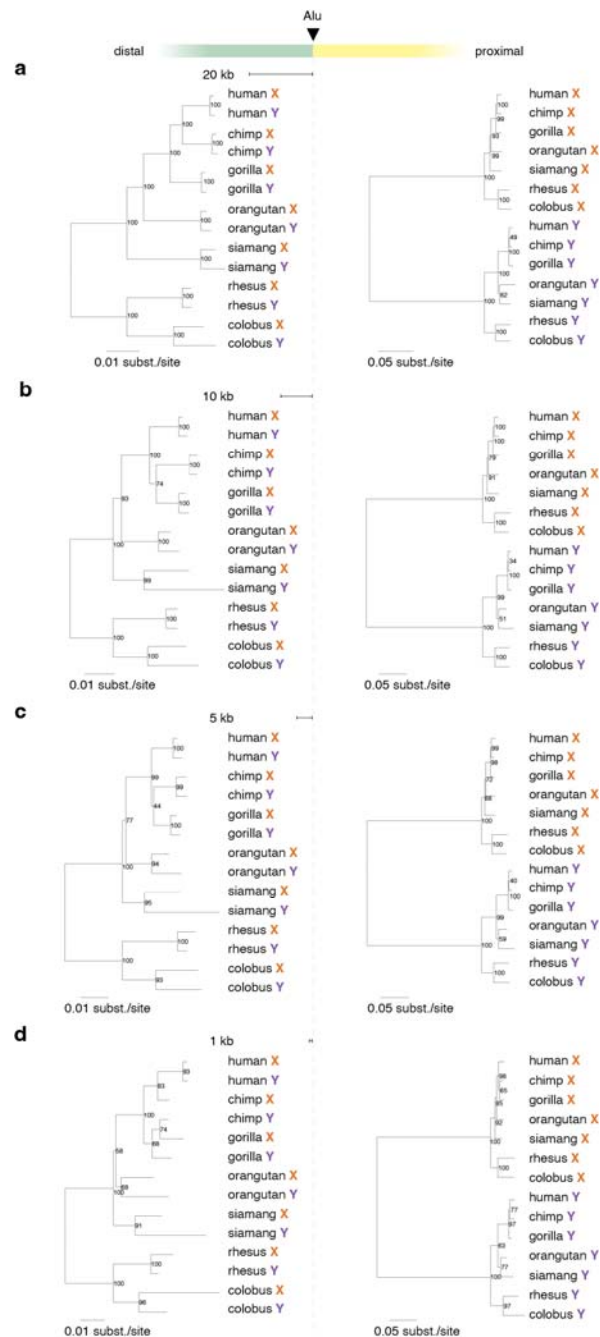
397 **Figure 3: Multiple alignment of primate sex chromosomes shows sharp change in X-Y identity**
398 **at the 1989 PAB.**

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

405

13
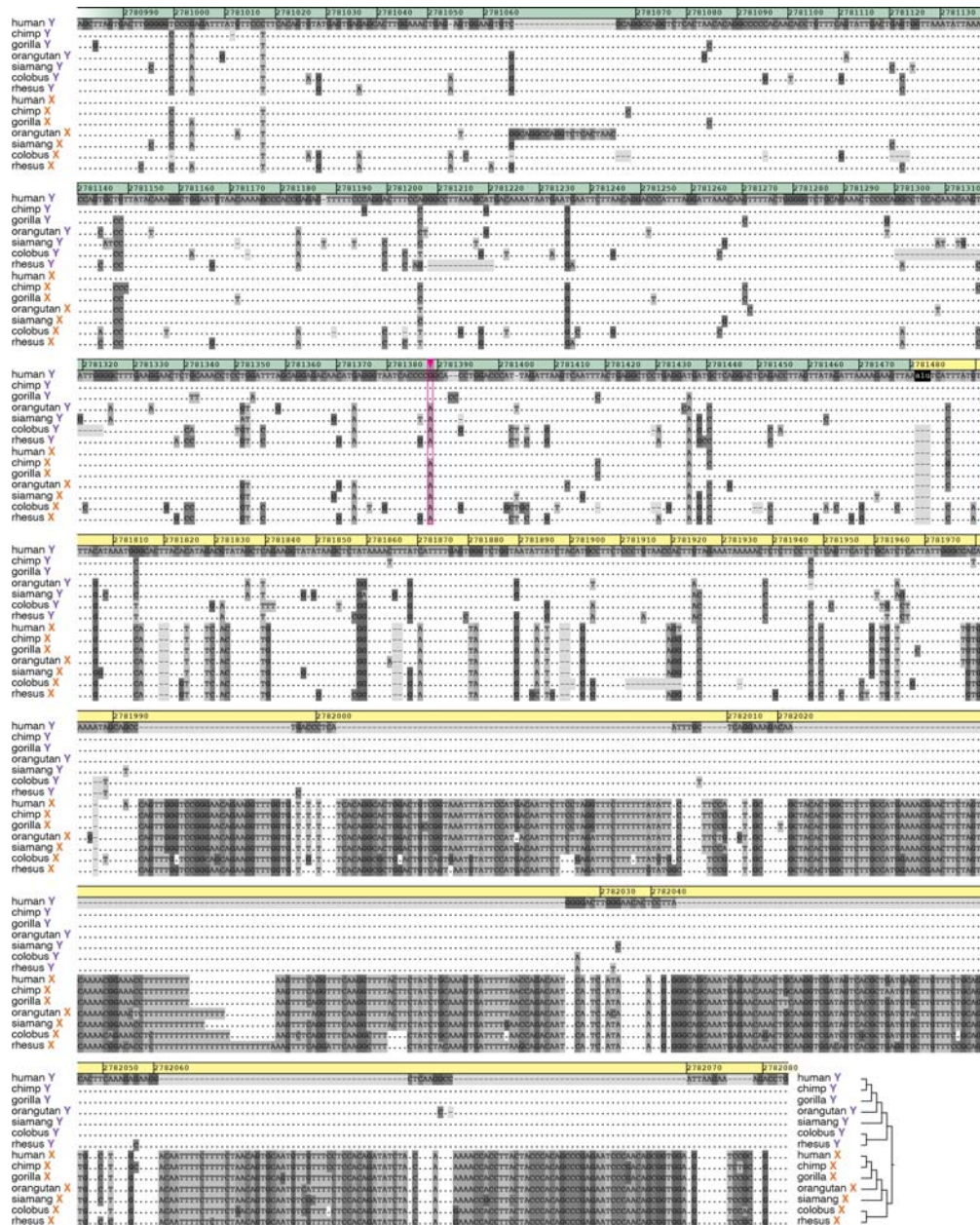
406   **Figure 4: Sex differences in MAF only occur within 10 kb of the 1989 PAB.**

407   Minor allele frequency (MAF) calculated from 1000 genomes data for the human PAR. Dark
408   green, sites <10 kb from the 1989 PAB; medium green, <500 kb from the 1989 PAB; light
409   green, all other sites. (a) Absolute sex difference in MAF plotted against hg38 location. 2023
410   and 1989 PAB indicated by dashed lines. (b-d) Male versus female MAF; (b) tight
411   correlation between male and female MAFs distal to the 2023 PAB; (c) similar correlation
412   at most positions between the 2023 and 1989 boundaries; except (d) in the 10 kb
413   immediately distal to the 1989 PAB.

414

415

14
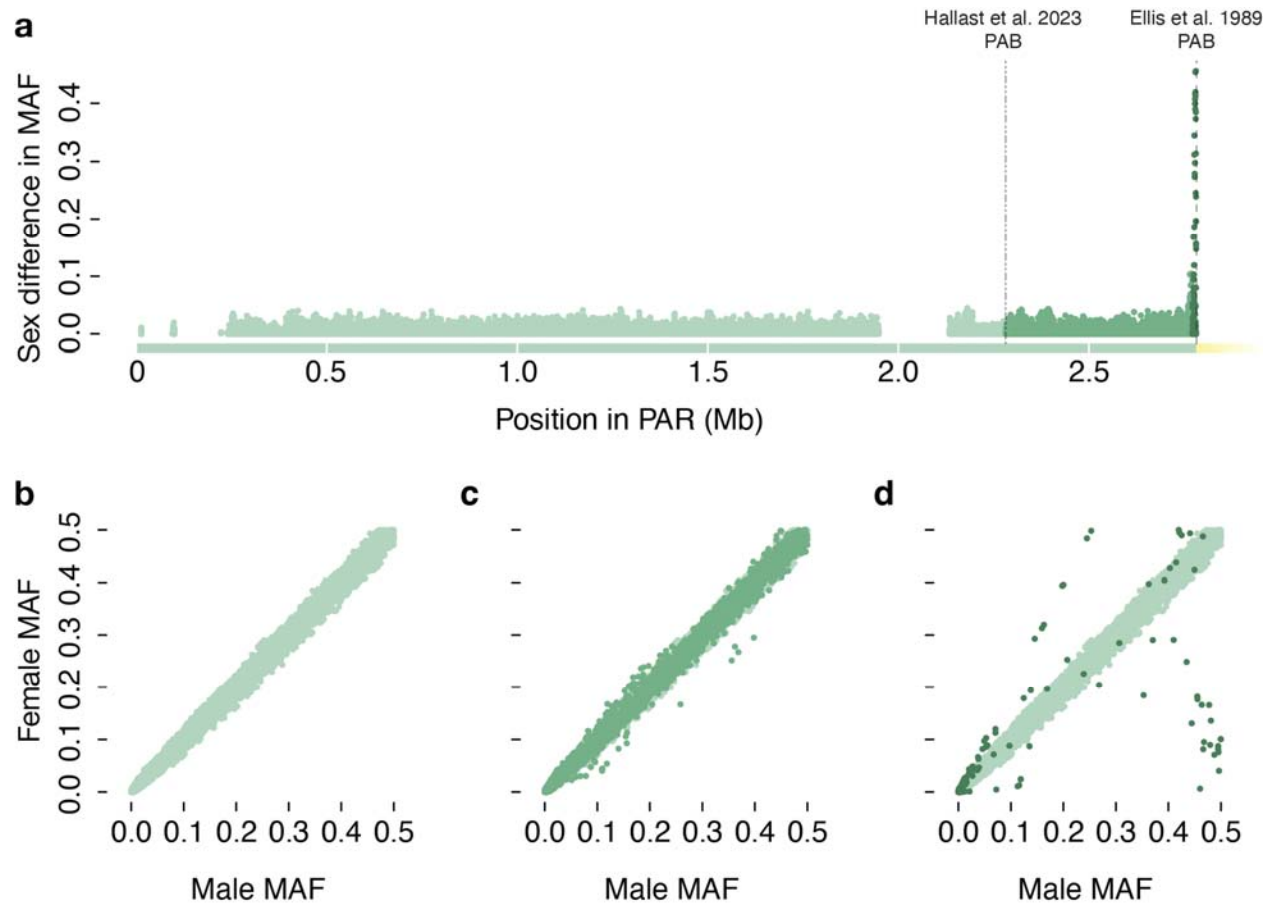
416 **Figure 5: Single-sperm sequencing reveals crossovers between 2023 and 1989 PABs.**

417 Summary of Sperm-seq results across 20 donors showing (a) average local recombination
418 rate between genotyped markers within PAR1, and (b) total genetic distance from the 1989
419 PAB. Medium green, region between 2023 PAB and 1989 PAB; light green, all other sites;
420 white tickmarks every 500 kb.

421

422

423 **Figure 6: Y chromosome sequences distal to the 1989 PAB are inconsistent with phylogenetic**
424 **tree of human NPY haplotypes.**

425 Results of LRT comparing single-clock versus local-clock models of evolution on the NPY
426 haplogroup trees in 20-kb sliding windows across multiple alignment of ten contiguous Y
427 chromosome sequences produced by Hallast and colleagues. Bonferroni corrected p-values
428 for LRT plotted versus hg38 position. 2023 and 1989 PAB indicated by dashed lines.
429 Windows distal to the 1989 PAB repeatedly reject the single-clock model.

16

430

431
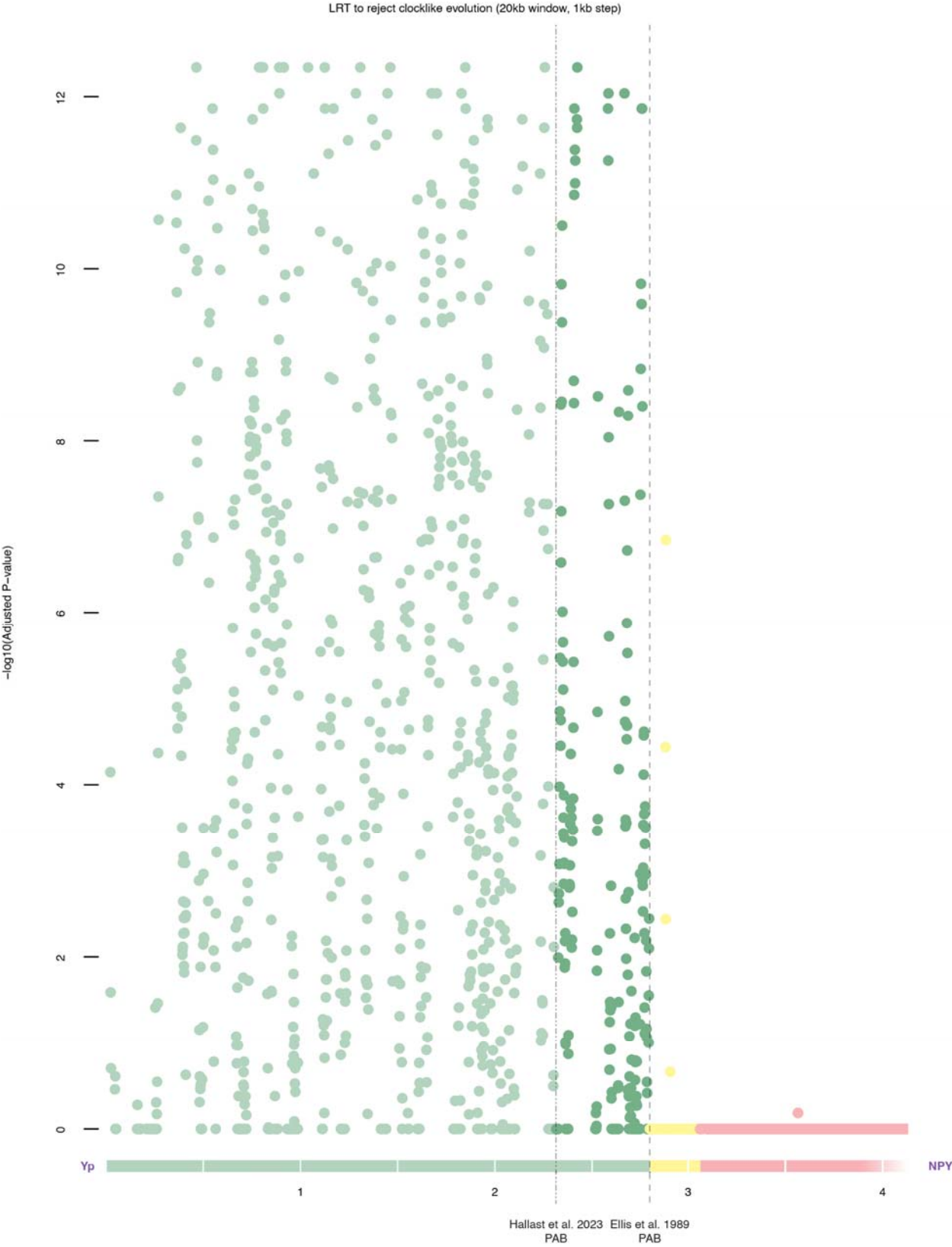
432 **Figure 7: Polymorphisms shared by X and Y chromosomes within 500 bases of the 1989 PAB.**

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

439

440

18
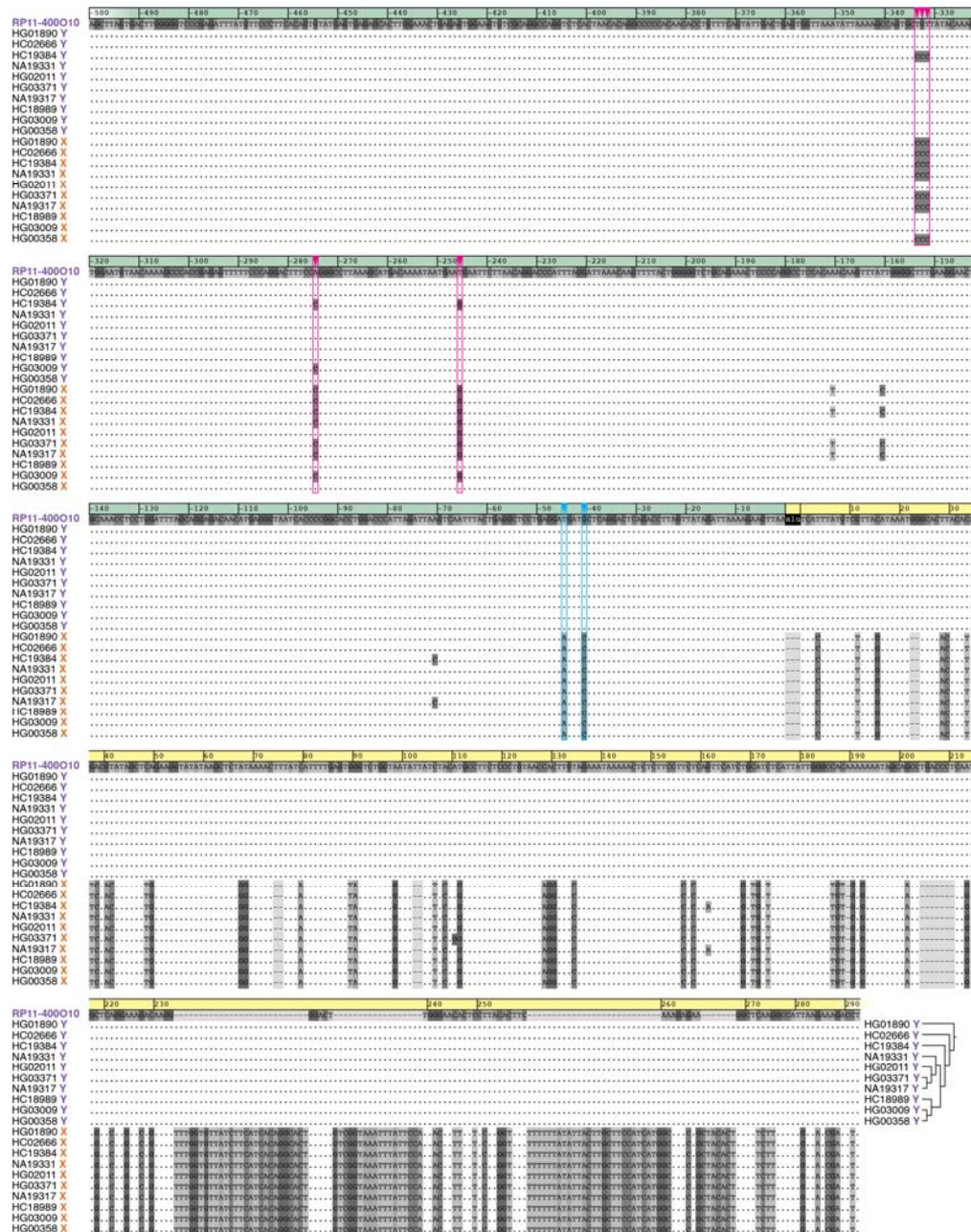
441   **Supplemental Figure 1: Single-sperm sequencing results.**

442   Summary of Sperm-seq results for 20 individual donors. Each plot shows local
443   recombination rate between genotyped markers within PAR1; number of genotyped sperm
444   indicated for each donor; medium green, region between the 2023 and 1989 PAB; light
445   green, all other sites; white tickmarks every 500 kb.


446

447
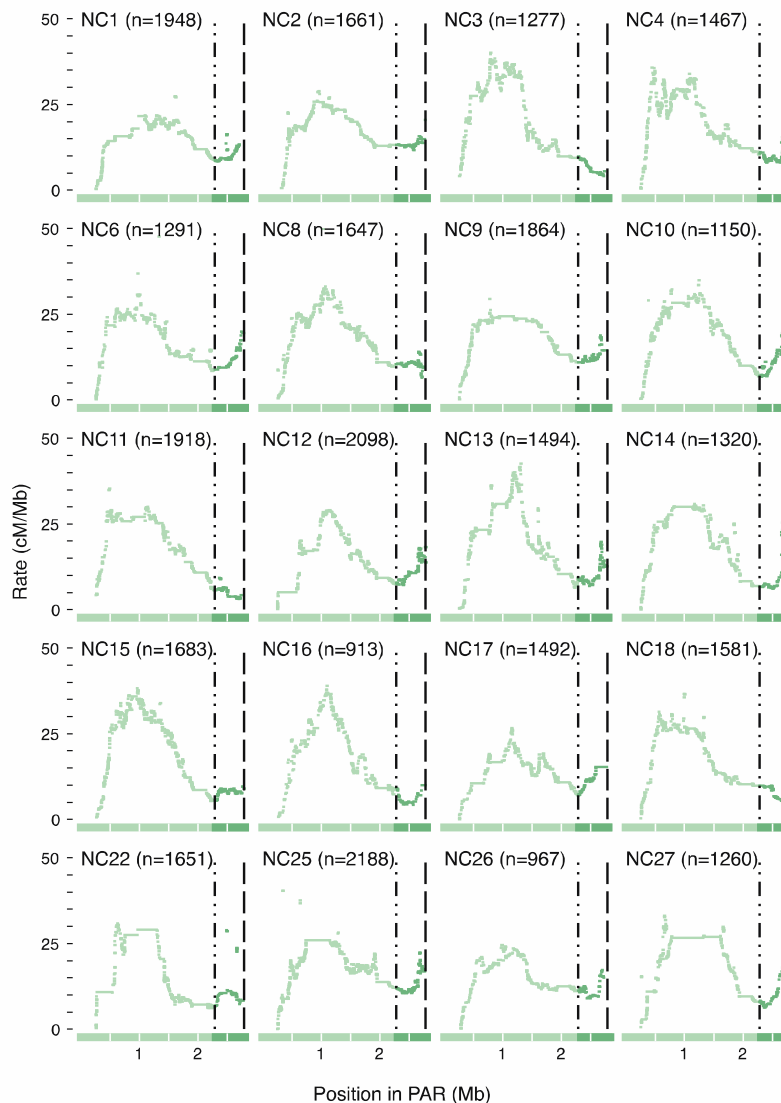
References

1. 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/kqn9.

2. Darlington, C.D., Haldane, J.B.S., and Koller, P.C.H. (1934). Possibility of Incomplete Sex Linkage in Mammals. Nature *133*, 417. https://doi.org/10/dn99bg.

3. Koller, P.Ch., and Darlington, C.D. (1934). The genetical and mechanical properties of the sex-chromosomes. J Genet *29*, 159–173. https://doi.org/10/dnrb4j.

4. 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/cbm4zw.

5. 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. https://doi.org/10/d2r8ps.

6. Goodfellow, P.J., Darling, S.M., Thomas, N.S., and Goodfellow, P.N. (1986). A Pseudoautosomal Gene in Man. Science (1979) *234*, 740–743. https://doi.org/10/cxdc37.

7. Kvaløy, K., Galvagni, F., and Brown, W.R.A. (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/d7k99g.

8. Ellis, N.A., Goodfellow, P.J., Pym, B., Smith, M., Palmer, M., Frischauf, 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/bb29qk.

9. Cooke, H.J., Brown, W.R.A., and Rappold, G.A. (1985). Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. Nature *317*, 687–692. https://doi.org/10/cs2fgz.

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/djgs7j.

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/fpnjpm.

12. Page, D.C., Bieker, K., Brown, L.G., Hinton, S., Leppert, M., Lalouel, J.-M., Lathrop, M., Nystrom-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/bfgxf7.

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/bqnnxv.

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/b83dk9.

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/dg7k56.

16. Ellis, N., Taylor, A., Bengtsson, B.O., Kidd, J., Rogers, J., and fellow, P.G. (1990). Population structure of the human pseudoautosomal boundary. Nature *344*, 663–665. https://doi.org/10/cnkn53.

17. Matise, T.C., Chen, F., Chen, W., Vega, F.M.D. La, 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/dztkjg.

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/d79d7r.

19. Nato, A.Q., Buyske, S., and Matise, T.C. (2018). The Rutgers map: A third-generation combined linkage-physical map of the human genome. Hum Genet Inst New Jersey Second Res Day.

20. 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/f6cxns.

21. 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.

22. 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. The American Journal of Human Genetics *66*, 557–566. https://doi.org/10/b4tntj.

521   23.   Lu, S., Zong, C., Fan, W., Yang, M., Li, J., Chapman, A.R., Zhu, P., Hu, X., Xu, L., Yan, L., et
522   al. (2012). Probing Meiotic Recombination and Aneuploidy of Single Sperm Cells by
523   Whole-Genome Sequencing. Science (1979) *338*, 1627–1630.
524   https://doi.org/10/f4gfjc.

525   24.   Löytynoja, A., and Goldman, N. (2005). An algorithm for progressive multiple
526   alignment of sequences with insertions. Proc Natl Acad Sci U S A *102*, 10557–10562.
527   https://doi.org/10/d94v78.

528   25.   Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O.
529   (2010). New algorithms and methods to estimate maximum-likelihood phylogenies:
530   assessing the performance of PhyML 3.0. Syst Biol *59*, 307–321.
531   https://doi.org/10/cg67gg.

532   26.   Byrska-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., Corvelo, A.,
533   Clarke, W.E., Musunuri, R., Nagulapalli, K., et al. (2022). High-coverage whole-genome
534   sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell
535   *185*, 3426–3440.e19. https://doi.org/10/gqq5vh.

536   27.   Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2.
537   Nat Methods *9*, 357–359. https://doi.org/10/gd2xzn.

538   28.   Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis,
539   G., Durbin, R., and Subgroup, 1000 Genome Project Data Processing (2009). The
540   Sequence Alignment/Map Format and SAMtools. Bioinformatics *25*, 2078–2079.
541   https://doi.org/10/ff6426.

542   29.   Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast
543   processing of NGS alignment formats. Bioinformatics *31*, 2032–2034.
544   https://doi.org/10/gfzsfw.

545   30.   Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-
546   read sequencing. Preprint at arXiv, https://doi.org/10/gtxrfb
547   https://doi.org/10/gtxrfb.

548   31.   Garrison, E., Kronenberg, Z.N., Dawson, E.T., Pedersen, B.S., and Prins, P. (2022). A
549   spectrum of free software tools for processing the VCF variant call format: vcflib, bio-
550   vcf, cyvcf2, hts-nim and slivar. PLoS Comput Biol *18*, e1009123.
551   https://doi.org/10/gqn62n.

552   32.   Bell, A.D., Mello, C.J., Nemesh, J., Brumbaugh, S.A., Wysoker, A., and McCarroll, S.A.
553   (2020). Insights into Variation in Meiosis from 31,228 Human Sperm Genomes.
554   Nature *583*, 259–264. https://doi.org/10/ggzrxc.

555   33.   Smit, A.F.A., Hubley, R., and Green, P. (2020). RepeatMasker. Preprint.

556   34.   Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences.
557   Nucleic Acids Res *27*, 573–580. https://doi.org/10/bnz59h.

558 35. Darling, A.E., Mau, B., and Perna, N.T. (2010). progressiveMauve: multiple genome
559   alignment with gene gain, loss and rearrangement. PLoS One *5*, e11147.
560   https://doi.org/10/df4rd4.

561 36. Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE:
562   Maximum Likelihood Phylogenetic Analysis Using Quartets and Parallel Computing.
563   Bioinformatics *18*, 502–504. https://doi.org/10/dcjhws.

564 37. R Core Team (2020). R: A Language and Environment for Statistical Computing (R
565   Foundation for Statistical Computing).

566 38. Okonechnikov, K., Golosova, O., Fursov, M., and team, U.G.E.N.E. (2012). Unipro
567   UGENE: a unified bioinformatics toolkit. Bioinformatics *28*, 1166–1167.
568   https://doi.org/10/ggbwmq.

569 39. Shen, M.R., Batzer, M.A., and Deininger, P.L. (1991). Evolution of the master Alu
570   gene(s). J Mol Evol *33*, 311–320. https://doi.org/10/bpdb3x.

571 40. Lahn, B.T., and Page (1999). Four Evolutionary Strata on the Human X Chromosome.
572   Science (1979) *286*, 964–967. https://doi.org/10/dxfhsf.

573 41. Bellott, D.W., Hughes, J.F., Skaletsky, H., Brown, L.G., Pyntikova, T., Cho, T.-J.,
574   Koutseva, N., Zaghlul, S., Graves, T., Rock, S., et al. (2014). Mammalian Y
575   Chromosomes Retain Widely Expressed Dosage-Sensitive Regulators. Nature *508*,
576   494–499. https://doi.org/10/f5zqhx.

577 42. Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: A Resource for
578   Timelines, Timetrees, and Divergence Times. Mol Biol Evol *34*, 1812–1819.
579   https://doi.org/10/gbhc3x.

580
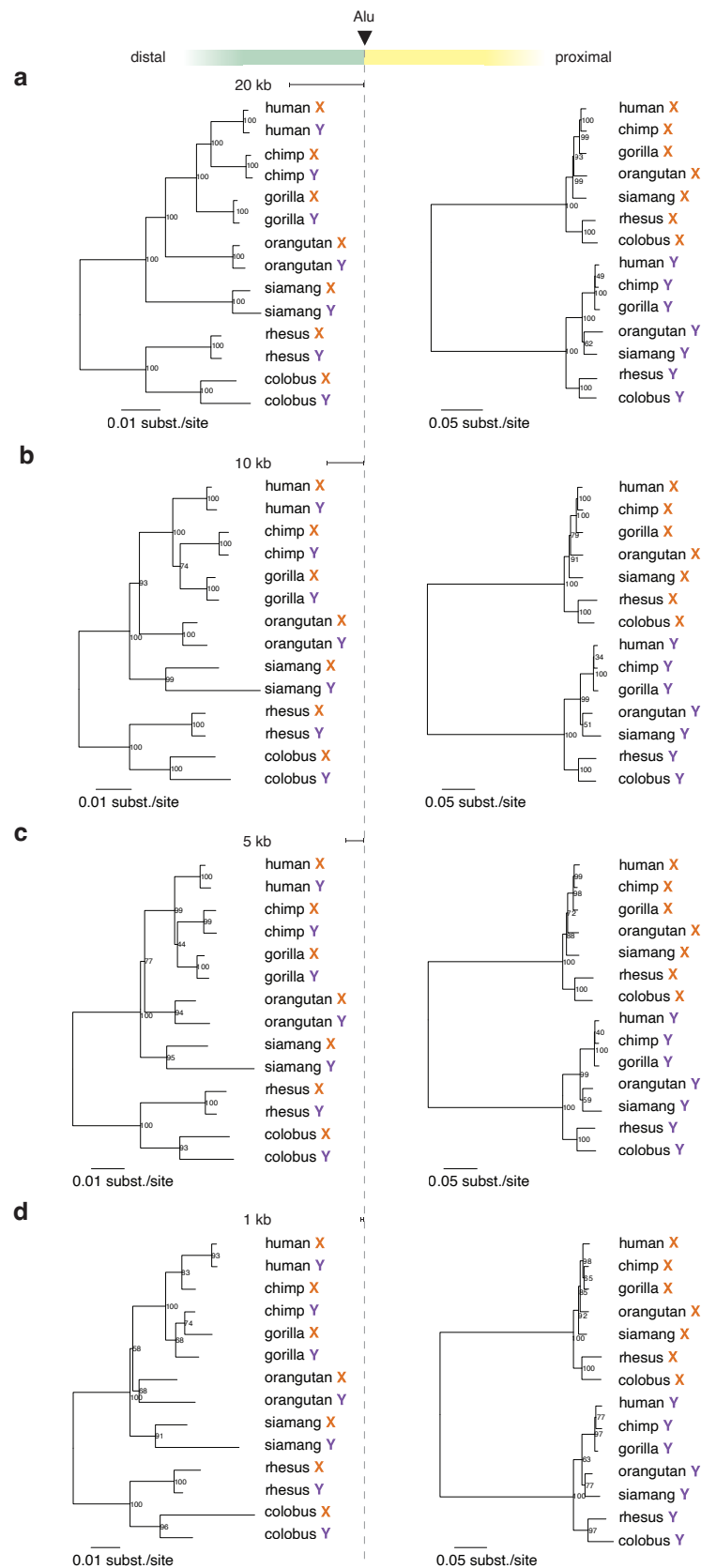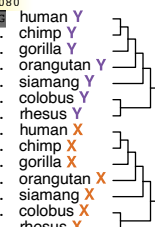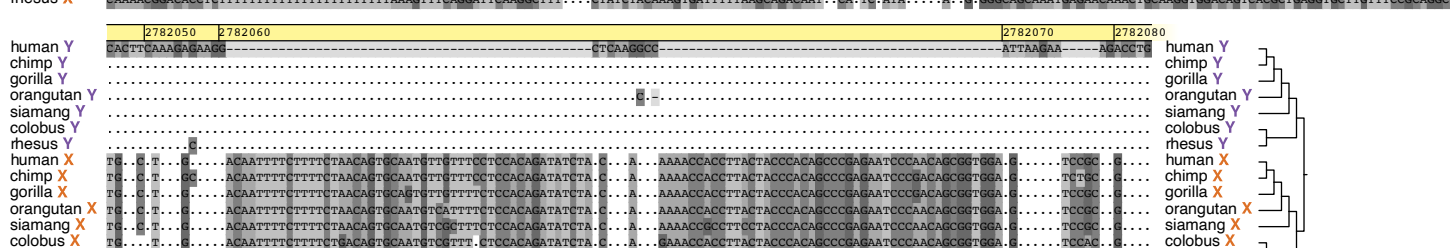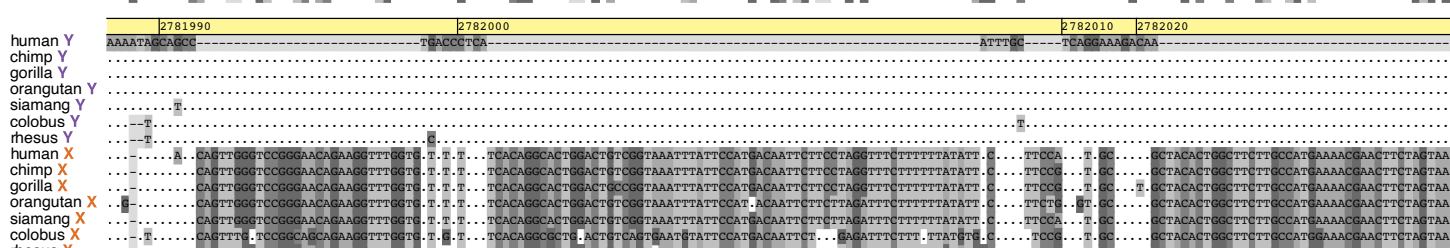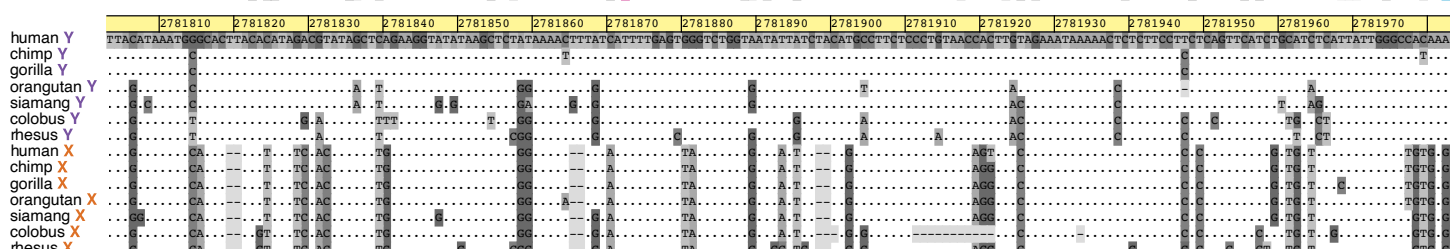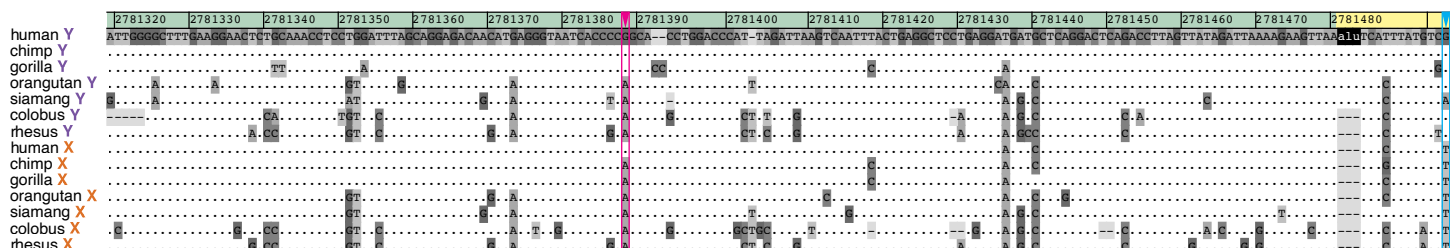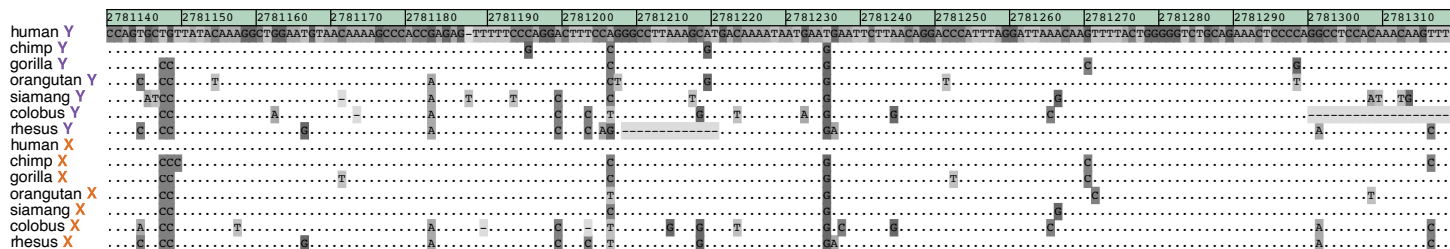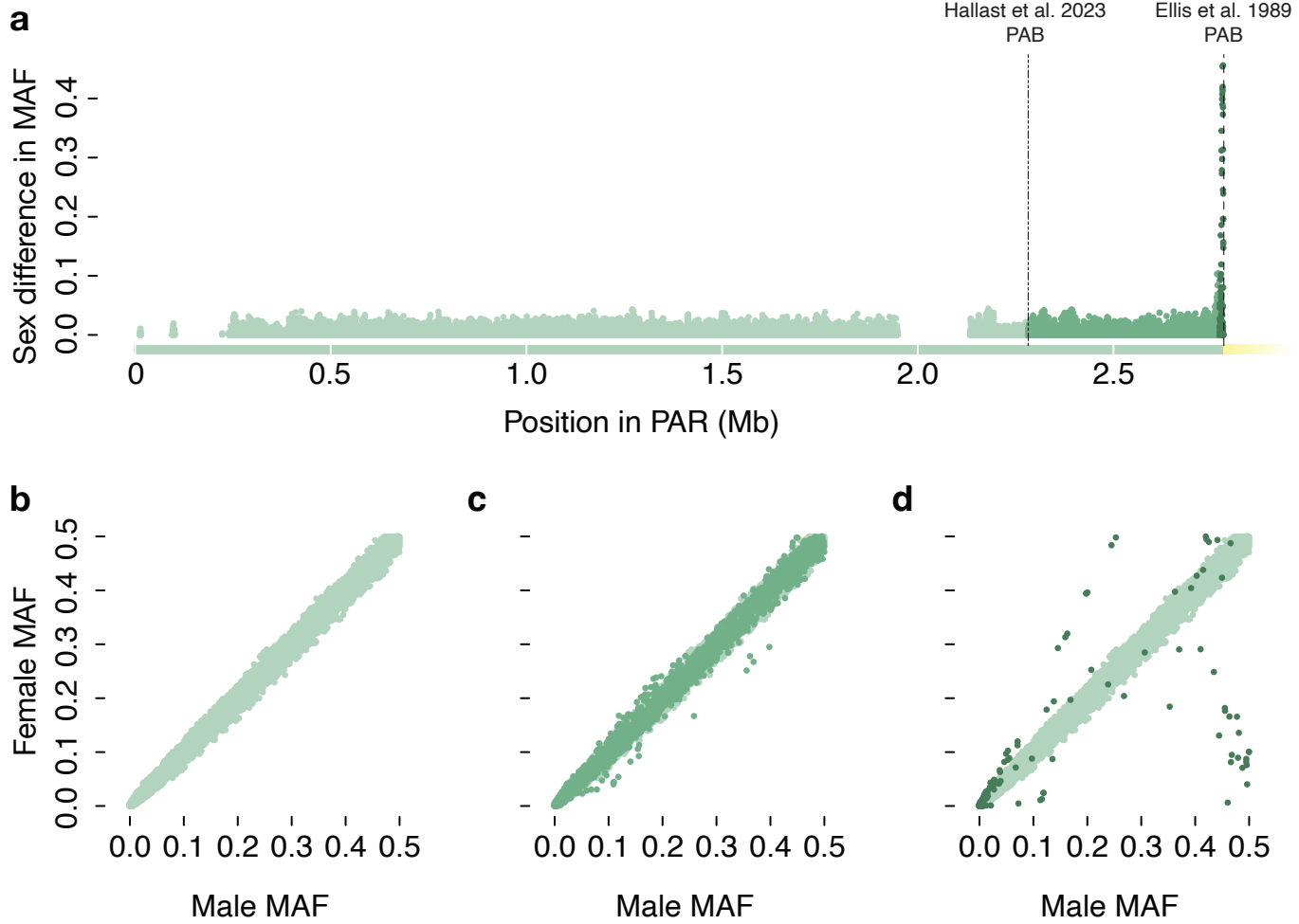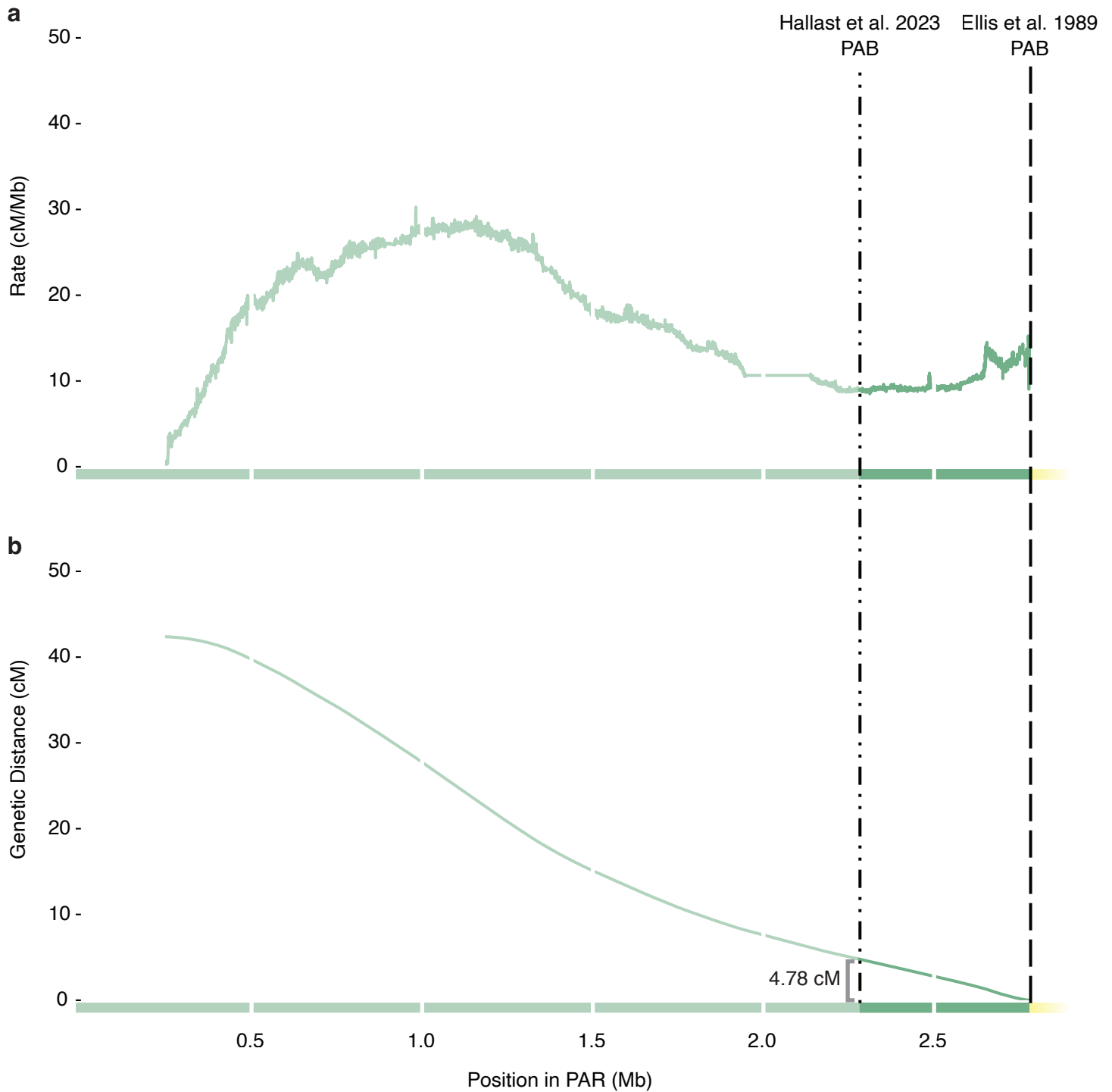
LRT to reject clocklike evolution (20kb window, 1kb step)