Genetic Diversity on the Human X Chromosome Does Not Support a Strict Pseudoautosomal Boundary

Daniel J. Cotter,*,1 Sarah M. Brotman,*,1 and Melissa A. Wilson Sayres*†
*School of Life Sciences and †Center for Evolution and Medicine, The Biodesign Institute, Arizona State University, Tempe, Arizona 85281

ABSTRACT Unlike the autosomes, recombination between the X chromosome and the Y chromosome is often thought to be constrained to two small pseudoautosomal regions (PARs) at the tips of each sex chromosome. PAR1 spans the first 2.7 Mb of the proximal arm of the human sex chromosomes, whereas the much smaller PAR2 encompasses the distal 320 kb of the long arm of each sex chromosome. In addition to PAR1 and PAR2, there is a human-specific X-transposed region that was duplicated from the X to the Y chromosome. The X-transposed region is often not excluded from X-specific analyses, unlike the PARs, because it is not thought to routinely recombine. Genetic diversity is expected to be higher in recombining regions than in nonrecombining regions because recombination reduces the effect of linked selection. In this study, we investigated patterns of genetic diversity in noncoding regions across the entire X chromosome of a global sample of 26 unrelated genetic females. We found that genetic diversity in PAR1 is significantly greater than in the nonrecombining regions (nonPARs). However, rather than an abrupt drop in diversity at the pseudoautosomal boundary, there is a gradual reduction in diversity from the recombining through the nonrecombining regions, suggesting that recombination between the human sex chromosomes spans across the currently defined pseudoautosomal boundary. A consequence of recombination spanning this boundary potentially includes increasing the rate of sex-linked disorders (e.g., de la Chapelle) and sex chromosome aneuploidies. In contrast, diversity in PAR2 is not significantly elevated compared to the nonPARs, suggesting that recombination is not obligatory in PAR2. Finally, diversity in the X-transposed region is higher than in the surrounding nonPARs, providing evidence that recombination may occur with some frequency between the X and Y chromosomes in the X-transposed region.

KEYWORDS genetics of sex; nucleotide diversity; pseudoautosomal region (PAR); X-transposed region (XTR); sex chromosome evolution; recombination

The human sex chromosomes, X and Y, were previously an indistinguishable pair of autosomes, but within the last 180–210 million years, the ancestral pair diverged into two distinct chromosomes of tremendously different gene content and function (Mikkelsen et al. 2007; Rens et al. 2007). The human sex chromosomes are composed of an older X-conserved region, shared across all therian (marsupial and eutherian) mammals (Watson et al. 1990; Glas et al. 1999), and a younger X- and Y-added region: an autosomal sequence that was translocated to the X and Y chromosomes in the common ancestor of eutherian mammals approximately 80–130 million years ago (Waters et al. 2001). The differentiation of the X and Y is hypothesized to have occurred after a series of Y-specific inversions that suppressed X-Y recombination (Lahn and Page 1999; Marnis and Galtier 2003; Lemaitre et al. 2009; Wilson and Makova 2009; Pandey et al. 2013). In the absence of homologous recombination, the Y chromosome has lost nearly 90% of the genes that were on the ancestral sex chromosomes (Skaletsky et al. 2003; Ross et al. 2005; Sayres and Makova 2013). Today, the human X and Y chromosomes share two pseudoautosomal regions (PARs) at the ends of the chromosomes that continue to undergo homologous X-Y recombination (Lahn and Page 1999). PAR1 spans the first 2.7 Mb of
the proximal arm of the human sex chromosomes (Ross et al. 2005) and contains genes from the ancient X- and Y-added region translocation. PAR1 is separated from the nonrecombining (nonPAR) regions on the Y chromosome by a Y-specific inversion that is hypothesized to suppress X-Y recombination at this pseudoautosomal boundary (Pandey et al. 2013). A functional copy of the XG gene spans the human pseudoautosomal boundary on the X chromosome (Yi et al. 2004) but is interrupted on the Y chromosome by a Y-specific inversion (Ellis et al. 1990). In contrast to this mechanism for PAR1 formation, the 320-kb human-specific PAR2 resulted from at least two duplications from the X chromosome to the terminal end of the Y chromosome (Charach et al. 2003).

Genes located in PAR1 have important functions in all humans. Although genes on one X chromosome in 46,XX individuals are silenced via a process called X-inactivation (Carrel and Willard 2005), which evolved in response to loss of homologous gene content on the Y chromosome (Wilson Sayres and Makova 2013), all 24 genes in PAR1 escape inactivation (Perry et al. 2001; Ross et al. 2005; Helena Mangs and Morris 2007) (Supplemental Material, Table S1). For example, one gene in PAR1, SHOX1, plays an important role in long bone growth and skeletal formation (Rao et al. 2001; Benito-Sanz et al. 2012; Tsuchiya et al. 2014). The consequences of SHOX1 disruption include short stature, skeletal deformities, Leri-Weill dyschondrosteosis (Ozaki et al. 2001), and a higher frequency of psychiatric disorders, including bipolar affective disorder (Flaquer et al. 2010).

The suggested function of the PARs is to assist in chromosome pairing and segregation (Kauppi et al. 2011). It has been proposed, in humans and in great apes, that crossover events are mandatory during male meiosis (Rouyer et al. 1986; Lien et al. 2000; Kauppi et al. 2012). Analyses of human sperm suggest that a deficiency in recombination in PAR1 is significantly correlated with the occurrence of nondisjunction and results in Klinefelter syndrome (47,XXY) (Shi et al. 2001). ASMT, another gene located in PAR1, is involved in the synthesis of melatonin and is thought to be connected with psychiatric disorders, including bipolar affective disorder (Flaquer et al. 2010).

The suggested function of the PARs is to assist in chromosome pairing and segregation (Kauppi et al. 2011). It has been proposed, in humans and in great apes, that crossover events are mandatory during male meiosis (Rouyer et al. 1986; Lien et al. 2000; Kauppi et al. 2012). Analyses of human sperm suggest that a deficiency in recombination in PAR1 is significantly correlated with the occurrence of nondisjunction and results in Klinefelter syndrome (47,XXY) (Shi et al. 2001). ASMT, another gene located in PAR1, is involved in the synthesis of melatonin and is thought to be connected with psychiatric disorders, including bipolar affective disorder (Flaquer et al. 2010).

The suggested function of the PARs is to assist in chromosome pairing and segregation (Kauppi et al. 2011). It has been proposed, in humans and in great apes, that crossover events are mandatory during male meiosis (Rouyer et al. 1986; Lien et al. 2000; Kauppi et al. 2012). Analyses of human sperm suggest that a deficiency in recombination in PAR1 is significantly correlated with the occurrence of nondisjunction and results in Klinefelter syndrome (47,XXY) (Shi et al. 2001). ASMT, another gene located in PAR1, is involved in the synthesis of melatonin and is thought to be connected with psychiatric disorders, including bipolar affective disorder (Flaquer et al. 2010).

Studies of human population genetic variation often compare diversity on the X chromosome with diversity on the autosomes to make inferences about sex-biased human demographic history (Hammer et al. 2008; Gottipati et al. 2011b; Arbiza et al. 2014). Typically, PAR1 and PAR2 are filtered out of these studies, at the defined pseudoautosomal boundaries, and the XTR is not filtered out. However, patterns of diversity across the entire human X chromosome, including transitions across the PARs and XTR, have not been investigated to justify these common practices. In this study, we investigate patterns of genetic diversity and divergence across the entire human X chromosome.

Materials and Methods

We analyzed X chromosomes from 26 unrelated (46,XX) individuals sequenced by CompleteGenomics (Drmanac et al. 2010) (Table S2). Sites were filtered, requiring that data be present (monomorphic or variable) in all 26 samples. Human-chimpanzee (hg19-panTro4), human-macaque (hg19-rheMac3), human-dog (hg19-canFam3), and human-mouse (hg19-mm10) alignments were extracted from the University of California Santa Cruz (UCSC) Genome Browser (Rosenbloom et al. 2015). We curated the human-chimpanzee and human-macaque alignments to filter out segments that included autosomal sequences aligning to the X chromosome (Table S3, Figure S1, and Figure S2). These alignments were visualized using Gmaj software (Blanchette et al. 2004). Additionally, we observed several regions across the X chromosome that exhibited heightened divergence between the human and chimpanzee or the human and macaque (Figure S3 and Figure S4). On further inspection, these regions often contain multi-copy gene families that could lead to mismapping (Table S3). Divergence estimates were similar with and without these
regions, and here we present results with these regions of high divergence near multicopy gene families excluded. Low-diversity (Dutheil et al. 2015) and ampliconic regions (Mueller et al. 2013; Nam et al. 2015) were filtered out of the data to avoid analyzing regions potentially affected by strong selective sweeps or difficult-to-align regions. Significant differences between PAR1 and nonPARs of the X chromosome, as well as significant differences between the XTR and nonXTRs of the X chromosome persist regardless of inclusion or exclusion of ampliconic and low-diversity regions (Figure 1, Table 1, Table S4, and Table S5).

We used Galaxy Tools (Blankenberg et al. 2011) to filter out regions that could cause potential sequence misalignments and regions defined by the UCSC Genome Browser (Rosenbloom et al. 2015) that may be subject to selection: Reference Sequence (RefSeq) database genes, simple repeats, and repetitive elements. We attempted to filter out noncoding regions near genes, but doing so would leave very little analyzable sequence in PAR1 and PAR2.

We measured the diversity between the sequences as $\pi$, the average pairwise nucleotide differences per site between all sequences in the sample:

$$\pi = \frac{2}{L} \sum_{i=1}^{k} p_i \sum_{j>i}^{k} p_j d_{ij}$$

where $L$ represents the number of called sites, $k$ represents the number of DNA sequences, $p_i$ and $p_j$ are the frequencies of the corresponding alleles $i$ and $j$, and $d_{ij}$ is the number of sites containing nucleotide differences. Diversity was calculated within each specific region (PAR1, PAR2, XTR, nonPARs with XTR, and nonPARs without XTR), as well as across sliding and nonoverlapping windows. We generated window-interval files across the human X chromosome with Galaxy Tools (Blankenberg et al. 2011) and conducted analysis in four sets of windows: (1) in a 1 Mb nonoverlapping window, (2) a 1 Mb window with 100 kb sliding start positions, (3) a 100 kb nonoverlapping window, and (4) a 100 kb window with 10 kb sliding start positions (Figure S5). We similarly calculated human-chimpanzee, human-macaque, human-dog, and human-mouse species divergence along the X chromosome in each of the four regions and in the same windows described previously. To normalize the data, $P$ values were divided by the observed divergence within the same interval.

Chromosome X was divided into windows that were permuted without replacement 10,000 times to assess significant differences between diversity in each region (PAR1, XTR, and PAR2) relative to nonPAR sequences. This analysis was repeated for uncorrected diversity and diversity corrected for human-chimpanzee, human-macaque, human-dog, and human-mouse divergence values. Empirical $P$-values were calculated by computing the number of times the difference between each pair of permuted sample regions was equal to or greater than the difference in observed diversity between each pair of regions. The negative correlation along the pseudoautosomal boundary was tested using linear regressions across 100 kb windows covering a total of 3 Mb for each regression (30 windows), shifting the window by 100 kb.
Table 1 Diversity across regions of the human X chromosome

| Region                  | Uncorrected π | π | Divergence | π | Divergence | π | Divergence | π | Divergence | π |
|-------------------------|---------------|---|------------|---|------------|---|------------|---|------------|---|
| nonPAR                  | 0.000602      | 0.009814 | 0.062865   | 0.049702 | 0.122744 | 0.234432 | 0.002566 | 0.305070 | 0.001972 |
| nonPARminus_XTR         | 0.000595      | 0.009782 | 0.062434   | 0.049512 | 0.121914 | 0.234372 | 0.002539 | 0.304460 | 0.001954 |
| PAR1                    | 0.011505      | 0.022643 | 0.066482   | 0.099892 | 0.015070 | 0.337717 | 0.004457 | NA       | 0.000000 |
| P vs. nonPAR            | 0.0000        | 0.3446   | 0.0070     | 0.0000   | 0.0077    | 0.0000   | NA        | NA       | NA       |
| P vs. nonPARminus_XTR   | 0.0000        | 0.4007   | 0.0000     | 0.0000   | 0.0000    | NA       | NA        | NA       | NA       |
| PAR2                    | 0.000678      | 0.008720 | 0.077794   | 0.040967 | 0.016559 | 0.218771 | 0.003101 | 0.257609 | 0.002633 |
| P vs. nonPAR            | 0.3094        | 0.5105   | 0.4804     | 0.4852   | 0.4852    | 0.5547   | 0.5547    | 0.5547   | 0.5547   |
| P vs. nonPARminus_XTR   | 0.3380        | 0.5268   | 0.4896     | 0.4814   | 0.4814    | 0.5824   | 0.5824    | 0.5824   | 0.5824   |
| XTR                     | 0.000747      | 0.010937 | 0.068256   | 0.056953 | 0.013108 | 0.245717 | 0.003038 | 0.336725 | 0.002217 |
| P vs. nonPAR            | 0.0004        | 0.0038   | 0.0223     | 0.0148   | 0.0007    | 0.0007   | 0.0007    | 0.0007   | 0.0007   |

Diversity, measured as the average number of pairwise differences per site (π) between the X chromosomes of 26 unrelated genetic females, in each region of the human X chromosome is presented first unnormalized for mutation-rate variation, then normalized using human-chimpanzee (hg19-panTro4) divergence, and then separately normalized for human-macaque (hg19-rheMac3), human-dog (hg19-canFam3), and human-mouse (hg19-mm10) divergence. The regions analyzed include the PAR1, PAR2, the XTR, and the nonPARs either including the XTR (nonPAR) or excluding the XTR (nonPARminus_XTR). The ampliconic and low-diversity regions have been filtered out. P-values from permutation tests with 10,000 replicates are shown between each recombining region and the nonPARs. All P-values are indicated in italics while bold corresponds to significant P-values.

Results

Human X-linked nucleotide diversity is high in PAR1 but not PAR2

We observe that uncorrected diversity is three times higher in PAR1 than in the nonPARs, whereas uncorrected diversity in PAR2 is not significantly greater than that in the nonPARs (Table 1, Figure 1, and Figure 3). We studied noncoding regions across the entire X chromosome, filtering out annotated genes, to minimize the effect of selection, but given their small sizes, we could not filter out regions far from genes in the PARs or XTR (see Materials and Methods). Ampliconic regions (Mueller et al. 2013; Nam et al. 2015), as well as regions of low diversity that are expected to have strong selective sweeps (Duthel et al. 2015), also were filtered out, which yielded the same result (Table S4). However, mutation-rate variation across the X chromosome may account for variable levels of diversity observed in the PARs and nonPARs. We normalized the nucleotide diversity to correct for mutation rate using pairwise divergence between humans and several different species: panTro4, rheMac3, canFam3, and mm10 (Table 1 and Figure S6). When we normalized with panTro4, the difference in diversity between PAR1 and nonPARs was not significant after filtering out the ampliconic regions, low-diversity regions, and the “not applicable” (NA) values. This could be a result of large variation in divergence across regions of the X chromosome between humans and chimpanzees, potentially owing to complex speciation events (Patterson et al. 2006). Given this phenomenon, we focus our interpretations on data that have been normalized using human-macaque divergence. Similar to the uncorrected diversity values, when we correct for mutation rate using macaque divergence values, we observe higher nucleotide diversity across humans in PAR1 and PAR2 relative to the nonPARs, with diversity being significantly higher in PAR1 than in nonPARs (with XTR removed) and not significantly different between PAR2 and nonPARs (Figure 1, Figure 3, and Table 1).

Curiously, human-chimpanzee and human-macaque divergence are quite high in PAR1 relative to the nonPARs in a pattern that does not reflect diversity (Figure 1 and Table 1). This result, predominantly, is due to high interspecies diversity in PAR1 and near the PAR boundary (Figure S3 and Figure S4). However, human-dog divergence roughly parallels uncorrected human diversity (Figure 1). Alignments between the human and the mouse in PAR1 are unavailable.

Further, significantly elevated diversity in PAR1 relative to the nonPARs cannot be attributed solely to mutation-rate variation across the X chromosome because the pattern remains after correction for divergence in each region (Figure 1 and Table 1). The pattern we observed is consistent with several processes, including selection reducing variation more at linked sites in the nonPARs than in PAR1 as a result of reduced rates of recombination in the nonPARs relative to the PARs or as a result of stronger drift in the nonPARs as a result of a smaller effective population size.

Data availability

The authors state that all data necessary for confirming the conclusions presented in this article are represented fully within the article. All codes used for this project can be found at https://github.com/WilsonSayresLab/PARdiversity.
That we do not observe significantly elevated diversity in PAR2 relative to the nonPARs is consistent with reports that PAR2 undergoes X-Y recombination less frequently than PAR1 (Flaquer et al. 2008) and supports assertions that in humans only one chiasma per chromosome is needed for proper segregation rather than one per chromosome arm (Fledel-Alon et al. 2009).

**Diversity is significantly higher in the XTR than in the nonPARs**

Curiously, in addition to elevated rates of diversity in the previously described PAR1 and PAR2, we also observed that diversity was significantly higher in the recent XTR than in the nonPARs (Table 1 and Figure 3). This increased diversity cannot be attributed to mismapping between the X and Y chromosome because we only analyzed individuals with two X chromosomes (see Materials and Methods). High diversity in the XTR contrasts with initial suggestions that there is no X-Y recombination in some humans populations in this region (Veerappa et al. 2013).

Given the large size of the nonPARs and the small size of the XTR, 5 Mb (Ross et al. 2005), one may wonder whether removing the XTR would make a difference to measured levels of diversity across the human X chromosome. The raw diversity of the nonPARs including the XTR, measured as $\pi$, is 0.000602, while the raw diversity of the nonPARs excluding the XTR is 0.000595 (Table 1). Removal of the XTR does decrease estimates of both diversity and divergence in the nonPARs. Although the XTR *de facto* may be removed with other filters, one should be cautious to include XTR regions because their inclusion in studies of X-specific diversity will affect inferences made when comparing X-linked and autosomal variation (Keinan and Reich 2010; Gottipati et al. 2011a; Wilson Sayres et al. 2014; Arbiza et al. 2014).

**Pseudoautosomal boundaries cannot be inferred from patterns of diversity**

Recombination between the X and Y chromosomes is expected to be suppressed at the pseudoautosomal boundary, where X-Y sequence homology diverges owing to a Y-specific inversion (Ellis et al. 1990; Yi et al. 2004; Pandey et al. 2013). If diversity correlates highly with recombination rate and X-Y recombination is strictly suppressed in the nonPARs after the pseudoautosomal boundary, then diversity is expected to drop sharply between PAR1 and the nonPARs. However, when we analyze patterns of human diversity in permuted windows across the X chromosome (see Materials and Methods), we do not observe an abrupt shift in the level of diversity between PAR1 and the nonPARs (Figure 2). The lack of an observable pseudoautosomal boundary based on diversity is clear whether small or large (100 kb or 1 Mb) or overlapping or nonoverlapping windows are used (Figure S5). In the approximately 3 Mb that span the pseudoautosomal boundary, we observe a significant negative correlation between distance from Xp and diversity. As we shift the window for the regression by 100 kb further from the start of PAR1, we observe that the negative correlations remain independently significant and continue past the boundary (Figure 2). We observe that the original linear relationship between distance from Xp and diversity has a significant negative coefficient of correlation ($R = -0.6681177$; $P = 0$) (Figure S7). The significant linear relationship ($P = 3.281 \times 10^{-10}$) that we observe in Figure 2 extends nearly twice the length of PAR1 and supports the observation that there is no clear, abrupt drop in nucleotide diversity across the pseudoautosomal boundary. To test the significance of this correlation, we conducted a permutation test, shuffling windows (of 100 kb) across the X chromosome and recomputing the series of linear regressions 10,000 times; then we computed the number of times a permuted X chromosome had a correlation that was as strong as or stronger than what we observed on the X chromosome (Figure S7). We found that the negative correlation between distance from the short arm of the X chromosome and diversity is significant and spans the pseudoautosomal boundary (see Materials and Methods; $P = 0$, permutation test).

The history of gene conversion between the sex chromosomes may contribute to the increased diversity levels...
Diversity along the X chromosome split by region. Genetic diversity (measured by $\pi$) is shown in box plots depicting the average diversity with error bars for the nonPARs, PAR1, XTR, and PAR2. The $p$-values from a permutation test with 10,000 replicates comparing the diversity of each region to the diversity of the nonPARs are shown.

Figure 3 Diversity (measured by $\pi$) is shown in box plots depicting the average diversity with error bars for the nonPARs, PAR1, XTR, and PAR2. The $p$-values from a permutation test with 10,000 replicates comparing the diversity of each region to the diversity of the nonPARs are shown.

(Trombetta et al. 2014) on the nonPAR side of the Y-specific inversion that marks the pseudoautosomal boundary. Human diversity uncorrected for divergence decreases from the proximal end of PAR1 through the pseudoautosomal boundary and well into the nonPAR. A sex-specific map of PAR1 found that male recombination is higher near the telomeres and decreases near the pseudoautosomal boundary, while, in contrast, the female recombination rate reported in the same study in PAR1 is fairly flat throughout the region and increases near the pseudoautosomal boundary (Hinch et al. 2014). Thus, genetic diversity uncorrected for divergence in PAR1 appears to correlate with the male recombination rate. Curiously, however, a previous study of recombination rate in PAR1 reported an increase in the female (but not the male) recombination rate near the proximal end of PAR1 (Henke et al. 1993). Thus, potentially, both male and female recombination rates contribute to the linear decrease in diversity observed in PAR1 from the proximal end of the X chromosome through the pseudoautosomal boundary. Although not yet mapped, when the data becomes available, it will be useful to compare patterns of diversity with sex-specific recombination maps across the entire X chromosome.

Discussion

We show that diversity is indeed higher in the pseudoautosomal regions and lower in the regions of the X chromosome that are not known to recombine in males (nonPARs). Diversity in PAR1 is significantly higher than in the nonPARs regardless of normalizing the diversity with divergence between human and either macaque or dog to correct for mutation rate (Table 1, Figure 1, and Figure 3). Diversity also was normalized with divergence from the mouse, but there is no alignment between human and mouse in PAR1 because of a different evolutionary origin in PAR1 and no common pseudoautosomal genes being shared between them (Gianfrancesco et al. 2001). We observed that diversity is lower in PAR2 than expected and is not significantly different from the nonPARs. We also showed that diversity is elevated in the XTR above other nonPARs, verifying recent observations that the region still may undergo homologous recombination between the X and Y chromosomes (Veerappa et al. 2013). Finally, when analyzing patterns of genetic diversity in windows across the human X chromosome, we found that there is no strict boundary, based solely on the levels of diversity, between the recombining and putatively nonrecombining regions, which could be attributed to the evolutionary shift in the pseudoautosomal boundary over time, extending PAR1 as a result of a PAR1 length polymorphism (Mensah et al. 2014). This also could suggest that nonhomologous recombination at the pseudoautosomal boundaries may be common.

Our observations of patterns of diversity across regions of the human X chromosome with variable levels of recombination are consistent with previous reports that diversity and divergence are correlated with recombination rate in humans across the genome (Hellmann et al. 2003) and explicitly in PAR1 (Bussell et al. 2006). Elevated levels of diversity in the XTR suggest that, consistent with a recent report (Veerappa et al. 2013), this region may frequently undergo X-Y recombination. Curiously, we did not find a significant elevation of diversity in PAR2, which, in agreement with its unusual evolution (Charchar et al. 2003), indicates that it rarely recombines between X and Y chromosomes during meiosis. Further, the lack of a clear differentiation in diversity between PAR1 and the nonPARs suggests that recombination suppression between the X and Y chromosomes is still an actively evolving process in humans, as in other species (Bergero and Charlesworth 2009). This is consistent with evidence that the position of the pseudoautosomal boundary varies across mammals (Raudsepp and Chowdhary 2008; Otto et al. 2011; Raudsepp et al. 2012; White et al. 2012). There is even evidence of polymorphism in the pseudoautosomal boundary in a pedigree analysis of a paternally inherited X chromosome in humans (Mensah et al. 2014). Recombination spanning the pseudoautosomal boundary may account for some cases of de la Chapelle syndrome (Schrander-Stumpel et al. 1994), in which an individual with two X chromosomes develops male gonads, and some portion of cases also have a copy of SRY (SRY sits immediately proximal to the pseudoautosomal boundary in humans). Further, it is possible that pseudoautosomal boundaries vary across populations, affecting recombination and contributing to nondisjunction of the sex chromosomes. Taken together with previous inferences about the variation in pseudoautosomal boundaries, our observations suggest that assumptions should not be made of a strict suppression of X-Y recombination at the proposed human pseudoautosomal boundary.
Acknowledgments

We thank the School of Life Sciences and the Biodis- 
goence Institute at Arizona State University for startup funding to M.A.W.S. This research was supported in part by funds from the School of Life Sciences Undergraduate Research Program (SOLUR) through the School of Life Sciences at Arizona State University, Tempe Campus. We thank the SOLUR Program and the College of Liberal Arts and Sciences Undergraduate Summer Enrichment program for support to S.M.B.

Literature Cited

Abbas, N., K. McElreavey, M. Leconiat, E. Vilain, F. Jaubert et al., 1993 Familial case of 46,XX male and 46,XX true hermaphro- 
dite associated with a paternal-derived SRY-bearing X chromo-
some. C. R. Acad. Sci. III 316: 375–383.

Arbiza, L., S. Gottipati, A. Siepel, and A. Keinan, 2014 Contrasting 
X-linked and autosomal diversity across 14 human populations. Am. J. Hum. Genet. 94: 827–844.

Benito-Sanz, S., M. Aza-Carmena, A. Rodríguez-Estevez, I. 
Rica-Etxebarria, R. Gracia et al., 2012 Identification of the 
first PAR1 deletion encompassing upstream SHOX enhancers in a family with idiopathic short stature. Eur. J. Hum. Genet. 
EJHG 20: 125–127.

Bergero, R., and D. Charlesworth, 2009 The evolution of re-
stricted recombination in sex chromosomes. Trends Ecol. Evol. 
24: 94–102.

Blankenberg, D., J. Taylor, A. Nekrutienko, and T. G. Team, 
2005 X-inactivation pro- 
le reveals 
expression and involvement in evolutionary formation of the 
present-day pseudoautosomal boundary of human sex chro-
mosomes. Hum. Mol. Genet. 5: 23–32.

Blankenberg, D., J. Taylor, A. Nekrutienko, and T. G. Team, 
2011 Making whole genome multiple alignments usable for 
biologists. Bioinformatics 27: 2426–2428.

Bussell, J. J., N. M. Pearson, R. Kanda, D. A. Filatov, and B. T. Lahn, 
2006 Human polymorphism and human-chimpanzee divergence in 
pseudoautosomal region correlate with local recombination rate. Gene 368: 94–100.

Carrel, L., and H. F. Willard, 2005 X-inactivation profile reveals 
extensive variability in X-linked gene expression in females. 
Nature 434: 400–404.

Charchar, F. J., M. Svartman, N. El-Mogharbel, M. Ventura, P. 
Kirby et al., 2003 Complex events in the evolution of the 
human pseudoautosomal region 2 (PAR2). Genome Res. 13: 
281–286.

Charlesworth, B., 2012 The effects of deleterious mutations on 
evolution at linked sites. Genetics 190: 5–22.

Dranac, R., A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns 
et al., 2010 Human genome sequencing using unchained base 
reads on self-assembling DNA nanoarrays. Science 327: 78–81.

Duthell, J. Y., K. Munch, K. Nam, T. Mailund, and M. H. Schierup, 
2015 Strong selective sweeps on the X chromosome in the 
human-chimpanzee ancestor explain its low divergence. PLoS 
Genet. 11: e1005451.

Ellis, N., A. Taylor, B. O. Bengtsson, J. Kidd, J. Rogers et al., 
1990 Population structure of the human pseudoautosomal 
boundary. Nature 344: 663–665.

Filatov, D. A., and D. T. Gerrard, 2003 High mutation rates in 
human and ape pseudoautosomal genes. Gene 317: 67–77.

Flaquer, A., G. A. Rappold, T. F. Wienker, and C. Fischer, 2008 The 
human pseudoautosomal regions: a review for genetic epidemi-
ologists. Eur. J. Hum. Genet. 16: 771–779.

Flaquer, A., R. A. Jamra, K. Etterer, G. O. Díaz, F. Rivas et al., 
2010 A new susceptibility locus for bipolar affective disorder in 
PAR1 on Xp22.3/Yp11.3. Am. J. Med. Genet. B Neuropsych- 
iatr. Genet. 153B: 1110–1114.

Fledel-Alon, A., D. J. Wilson, K. Broman, X. Wen, C. Ober et al., 
2009 Broad-scale recombination patterns underlying proper 
disjunction in humans. PLoS Genet. 5: e1000658.

Fukagawa, T., Y. Nakamura, K. Okumura, M. Nomari, A. Ando 
et al., 1996 Human pseudoautosomal boundary-like sequences: 
expression and involvement in evolutionary formation of the 
present-day pseudoautosomal boundary of human sex chro-
mosomes. Hum. Mol. Genet. 5: 23–32.

Gianfrancesco, F., R. Sanges, T. Esposito, S. Tempesta, E. Rao et al., 
2001 Differential divergence of three human pseudoautosomal 
genes and their mouse homologs: implications for sex chromo-
some evolution. Genome Res. 11: 2095–2100.

Glas, R., J. A. M. Graves, R. Toder, M. Ferguson-Smith, and P. C. 
O’Brien, 1999 Cross-species chromosome painting between 
human and marsupial directly demonstrates the ancient region of the 
mammalian X. Mamm. Genome 10: 1115–1116.

Gottipati, S., L. Arbiza, A. Siepel, A. G. Clark, and A. Keinan, 
2011a Contrasting human X-linked and autosomal variation in 
population-scale whole genome sequencing. Am. J. Hum. 
Genet. 94: 828–830.

Hellmann, I., I. Ebersberger, S. E. Ptak, S. Pääbo, and M. Przeworski, 
2003 A neutral explanation for the correlation of diversity with 
recombination rates in humans. Am. J. Hum. Genet. 72: 1527–1535.

Henke, A., C. Fischer, and G. A. Rappold, 1993 Genetic map of the 
human pseudoautosomal region reveals a high rate of recombi-
nation in female meiosis at the Xp telomere. Genomics 18: 478–485.

Hinch, A. G., N. Altemose, N. Noor, P. Donnelly, and S. R. Myers, 
2014 Recombination in the human pseudoautosomal region 
PAR1. PLoS Genet. 10: e1004503.

Huang, S.-W., R. Friedman, N. Yu, A. Yu, and W.-H. Li, 2005 How 
strong is the mutagenicity of recombination in mammals? Mol. 
Biol. Evol. 22: 426–431.

Kauppi, L., M. Barchi, F. Baudat, P. J. Romanienko, S. Keeney et al., 
2011 Distinct properties of the XY pseudoautosomal region 
crucial for male meiosis. Science 331: 916–920.

Keinan, A., and D. Reich, 2010 Can a sex-biased human demog-
graphy account for the reduced effective population size of chro-
mosome X in non-Africans? Mol. Biol. Evol. 27: 2321–2321.

Lahn, B. T., and D. C. Page, 1999 Four evolutionary strata on the 
human X chromosome. Science 286: 964–967.

Lemaitre, C., M. D. V. Braga, C. Gautier, M.-F. Sagot, E. Tannier 
et al., 2009 Footprints of inversions at present and past pseudo-
autosomal boundaries in human sex chromosomes. Genome 
Biol. Evol. 1: 56–66.

Lien, S., J. Szyda, B. Schechinger, G. Rappold, and N. Armheim, 
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.
Marais, G., and N. Galtier, 2003 Sex chromosomes: how X-Y recombination stops. Curr. Biol. 13: R641–R643.
Mensah, M. A., M. S. Hestand, M. H. D. Larmeurse, M. Isrie, N. Vanderheyden et al., 2014 Pseudoautosomal region 1 length polymorphism in the human population. PLoS Genet. 10: e1004578.
Mikkelsen, T. S., M. J. Wakefield, B. Aken, C. T. Amemiya, J. L. Chang et al., 2007 Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature 447: 167–177.
Mueller, J. L., H. Skaletsky, L. G. Brown, S. Zaghhlul, S. Rock et al., 2013 Independent specialization of the human and mouse X chromosomes for the male germ line. Nat. Genet. 45: 1083–1087.
Nam, K., K. Munch, A. Hobolth, J. Y. Dutheil, K. R. Veeramah et al., 2015 Extreme selective sweeps independently targeted the X chromosomes of the great apes. Proc. Natl. Acad. Sci. USA 112: 6413–6418.
Otto, S. P., J. R. Pannell, C. L. Peichel, T.-L. Ashman, D. Charlesworth et al., 2011 About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet. 27: 358–367.
Page, D. C., A. de la Chapelle, and J. Weissenbach, 1985 Chromosome Y-specific DNA in related human XX males. Nature 315: 224–226.
Pandey, R. S., M. A. Wilson, Sayres, and R. K. Azad, 2013 Detecting evolutionary strata on the human X chromosome in the absence of gametologous Y-linked sequences. Genome Biol. Evol. 5: 1863–1871.
Patterson, N., D. J. Richter, S. Gnerre, E. S. Lander, and D. Reich, 2006 Genetic evidence for complex speciation of humans and chimpanzees. Nature 441: 1103–1108.
Perry, J., and A. Ashworth, 1999 Evolutionary rate of a gene affected by chromosomal position. Curr. Biol. 9: 987–989.
Perry, J., S. Palmer, A. Gabriel, and A. Ashworth, 2001 A short pseudoautosomal region in laboratory mice. Genome Res. 11: 1826–1832.
Rao, E., B. Weiss, M. Fukami, A. Rump, B. Niesler et al., 1997 Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. Nat. Genet. 16: 54–63.
Rao, E., R. J. Blaschke, A. Marchini, B. Niesler, M. Burnett et al., 2001 The Leri-Weill and Turner syndrome homeobox gene SHOX encodes a cell-type specific transcriptional activator. Hum. Mol. Genet. 10: 3083–3091.
Raussepp, T., and B. P. Chowdhary, 2008 The horse pseudoautosomal region (PAR): characterization and comparison with the human, chimp and mouse PARs. Cytogenet. Genome Res. 121: 102–109.
Raussepp, T., P. J. Das, F. Avila, and B. P. Chowdhary, 2012 The pseudoautosomal region and sex chromosome aneuploidies in domestic species. Sex Dev. 6: 72–83.
R Core Team, 2015 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rens, W., P. C. O’Brien, F. Grützner, O. Clarke, D. Graphodatskaya et al., 2007 The multiple sex chromosomes of platypus and echidna are not completely identical and several share homology with the avian Z. Genome Biol. 8: R243.
Rosenbloom, K. R., J. Armstrong, G. P. Barber, J. Casper, H. Clawson et al., 2015 The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 43: D670–D681.
Ross, M. T., D. V. Graffham, A. J. Coffey, S. Scherer, K. McLaY et al., 2005 The DNA sequence of the human X chromosome. Nature 434: 325–337.
Rosser, Z. H., P. Balaresque, and M. A. Jobling, 2009 Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot. Am. J. Hum. Genet. 85: 130–134.
Rouyer, F., M.-C. Simmler, C. Johnsson, G. Vergnaud, H. J. Cooke et al., 1986 A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. Nature 319: 291–295.
Schneider-Stumpep, C., M. Havenith, E. V. Linden, W. Maertzdorf, J. Offermanns et al., 1994 De la Chapelle dysplasia (atelosteogenesis type II): case report and review of the literature [corrected]. Clin. Dysmorphol. 3: 318–327.
Shi, Q., E. Spriggs, L. L. Field, A. Rademaker, E. Ko et al., 2002 Absence of age effect on meiotic recombination between human X and Y chromosomes. Am. J. Hum. Genet. 71: 254–261.
Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier et al., 2003 The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423: 825–837.
Trombetta, B., D. Sellitto, R. Scozzari, and F. Cruciani, 2014 Inter- and intraspecies phylogenetic analyses reveal extensive X-Y gene conversion in the evolution of gametologous sequences of human sex chromosomes. Mol. Biol. Evol. 31: 2108–2123.
Tsuchiya, T., M. Shibata, H. Numabe, T. Jinno, K. Nakabayashi et al., 2014 Compound heterozygous deletions in pseudoautosomal region 1 in an infant with mild manifestations of langer mesomelic dysplasia. Am. J. Med. Genet. A 164: 505–510.
Veerappa, A. M., P. Padakannaya, and N. B. Ramachandra, 2013 Copy number variation-based polymorphism in a new pseudoautosomal region 3 (PAR3) of a human X-chromosome-transposed region (XTR) in the Y chromosome. Funct. Integr. Genomics 13: 285–293.
Vicoso, B., and B. Charlesworth, 2006 Evolution on the X chromosome: unusual patterns and processes. Nat. Rev. Genet. 7: 645–653.
Waters, P. D., B. Duffy, C. J. Frost, M. L. Delbridge, and J. A. Graves, 2001 The human Y chromosome derives largely from a single autosomal region added to the sex chromosomes 80–130 million years ago. Cytogenet. Cell Genet. 92: 74–79.
Watson, J. M., J. A. Spencer, A. D. Riggs, and J. A. Graves, 1990 The X chromosome of monotremes shares a highly conserved region with the eutherian and marsupial X chromosomes despite the absence of X chromosome inactivation. Proc. Natl. Acad. Sci. USA 87: 7125–7129.
White, M. A., A. Ikeda, and B. A. Payseur, 2012 A pronounced evolutionary shift of the pseudoautosomal region boundary in house mice. Mamm. Genome 23: 454–466.
Wilson, M. A., and K. D. Makova, 2009 Evolution and survival on eutherian sex chromosomes. PLoS Genet. 5: e1000568.
Wilson Sayres, M. A., and K. D. Makova, 2013 Gene survival and death on the human Y chromosome. Mol. Biol. Evol. 30: 781–787.
Wilson Sayres, M. A., K. E. Lohmueller, and R. Nielsen, 2014 Natural selection reduced diversity on human Y chromosomes. PLoS Genet. 10: e1004064.
Yi, S., T. J. Summers, N. M. Pearson, and W.-H. Li, 2004 Recombination has little effect on the rate of sequence divergence in pseudoautosomal boundary 1 among humans and great apes. Genome Res. 14: 37–43.

Communicating editor: J. M. Akey
Genetic Diversity on the Human X Chromosome Does Not Support a Strict Pseudoautosomal Boundary

Daniel J. Cotter, Sarah M. Brotman, and Melissa A. Wilson Sayres
Figure S1. Divergence along the X chromosome. Divergence is shown in 100kb overlapping windows across the human X chromosome which includes the pseudoautosomal region 1 (PAR1), shown in red, the non-pseudoautosomal region, shown in black, the X-transposed region (XTR), shown in blue, and the pseudoautosomal region 2 (PAR2), shown in red, for A) human-chimpanzee; B) human-macaque; C) human-dog and D) human-mouse.
Figure S2. Human-chimpanzee divergence along chromosome X. Divergence values between hg19-panTro4 are shown, computed in 100kb windows with 10kb sliding start positions for A) not filtered for high divergence regions; and, B) with high divergence regions filtered out. 

1
Figure S3. Human-macaque divergence along chromosome X. Divergence values between hg19-rheMac3 are shown, computed in 100kb windows with 10kb sliding start positions, for A) not filtered for high divergence regions; and, B) with high divergence regions filtered out \(^1\).
Figure S4. Diversity uncorrected for divergence in variable windows. Diversity is depicted in A) 100kb nonoverlapping, B) 100kb overlapping, C) 1Mb nonoverlapping, and D) 1Mb overlapping windows.\(^1\)
Figure S5. Linear correlation of distance from Xpter and diversity on the X chromosome. Correlation coefficients were calculated for linear regressions through the first 3 Mb of 10,000 random permutations of the 100kb non-overlapping diversity data. These values were compared against the value for the original data.
Figure S6. Coverage by region for human-chimpanzee alignments. Coverage is plotted with gmaj^2, for A) PAR1, B) PAR2, and C) XTR.
Figure S7. Coverage by region for human-rhesus macaque data. Coverage is plotted with gmaj^2, for A) PAR1, B) PAR2, and C) XTR.
**Table S1. Genes in recombining regions of the X chromosome.** Genes found in PAR1, PAR2, and XTR regions.

| Region | Gene     | RefSeq ID    | Summary from NCBI Gene                                      | Citation |
|--------|----------|--------------|-------------------------------------------------------------|----------|
| PAR1   | PLCXD1   | NR_028057.1  | Terminal protein-coding gene.                               | 3        |
|        | GTPBP6   | NM_012227.3  | Encodes GTP binding protein.                                | 4        |
|        | LINC00685| NR_027231.1  | Long intergenic non-protein coding RNA.                     | 5        |
|        | PPP2R3B  | NM_013239.4  | Ser/Thr phosphatase implicated in the negative control of cell growth and division. | 6        |
|        | SHOX     | NM_000451.3  | Homeobox family. Defects are linked with idiopathic growth retardation and the short stature phenotype of Turner syndrome. | 7        |
|        | CRLF2    | NM_001012288.2| Member of the type I cytokine receptor family. Helps control cell proliferation. | 8        |
|        | MIR3690  | NR_037461.1  | Non-coding RNAs involved in post-transcriptional regulation of gene expression. | 9        |
|        | IL3RA    | NM_002183.3  | Encodes an interleukin 3 specific subunit of a heterodimeric cytokine receptor. | 10       |
|        | SLC25A6  | NM_001636.3  | Serves as a gated pore that translocates ADP and ATP between the cytoplasm and mitochondrial matrix. | 11       |
|        | LINC00106| NR_130733.1  | Long intergenic non-protein coding RNA.                     | 5        |
|        | ASMT     | NM_004043.2  | Belongs to methyltransferase superfamily. Catalyzes the final reaction in the synthesis of melatonin. | 12       |
|        | P2RY8    | NM_178129.4  | G-protein coupled receptor.                                 | 13       |
|        | AKAP17A  | NM_005088.2  | Protein kinase A anchoring protein. Part of the spliceosome complex. | 14       |
|        | DHRSX    | NM_145177.2  | Dehydrogenase/reductase (SDR family) X-linked.              | 15       |
|        | ZBED1    | NM_004729.3  | Zinc finger. Functions as a transcription factor.           | 16       |
|        | MIR6089  | NR_106737.1  | Short non-coding RNAs involved in post-transcriptional regulation of | 9        |
| Gene   | Accession    | Description                                                                                     | Page |
|--------|--------------|-------------------------------------------------------------------------------------------------|------|
| CD99   | NM_002414.2  | Cell surface glycoprotein.                                                                       | 17   |
| LINC00102 | NR_037842.1 | Long intergenic non-protein coding RNA.                                                          | 5    |
| XG     | NM_175569.2  | XG blood group antigen and spans the pseudoautosomal boundary.                                  | 18   |
| XGY2   | NR_003254.1  | XG pseudogene, y-linked 2                                                                       | 19   |
| XTR    | TGIF2LX      | Member of the TALE/TGIF homeobox family of transcription factors.                               | 20   |
| PABPC5 | NM_080832.2  | Binds to the polyA tail of eukaryotic mRNAs.                                                    | 21   |
| PCDH11X| NM_001168361.1 | Belongs to the protocadherin gene family. Thought to play a role in cell-cell recognition. | 22   |
| MIR4454| NR_039659.1  | Affects the stability and translation of mRNAs. Involved in post-transcriptional regulation of gene expression. | 9    |
| NAP1L3 | NM_004538.5  | Has no introns and is a member of the nucleosome assembly protein family.                       | 23   |
| FAM133A| NM_001171111.1 | Family with sequence similarity 133.                                                               | 5    |
| PAR2   | SPRY3        | Sprouty RTK signaling antagonist 3.                                                              | 24   |
| VAMP7  | NR_033715.1  | Member of the soluble N-ethylmaleimide-sensitive attachment protein receptor (SNARE) family. Involved in the fusion of transport vesicles to their target membranes. | 25   |
| IL9R   | NM_002186.2  | Cytokine receptor that mediates the biological effects of interleukin 9.                         | 26   |
| DDX11L16 | NR_110561.1 | DEAD/H (Asp-Glu_Ala-Asp/His) box helicase 11 like 16.                                              | 27   |
Table S2. Diversity in the ampliconic and low diversity regions. The table shows the pi value computed for the nonPAR region of the X chromosome. This compares the nonPAR, nonPAR with ampliconic regions filtered out, nonPAR with low diversity regions filtered out, and nonPAR with both sets of regions filtered out. It also shows the pi values for each ampliconic and each low diversity region.

| Region                  | Chr X position                  | Pi           |
|-------------------------|---------------------------------|--------------|
| nonPAR                  | 27000000-154940559              | 0.00051179   |
| Ampliconic region 1     | 48202745-48292983               | 0.00000447   |
| Ampliconic region 2     | 48976199-49062381               | 0.00000678   |
| Ampliconic region 3     | 51395467-51492862               | 0.00000235   |
| Ampliconic region 4     | 51775560-51966529               | 0.00000402   |
| Ampliconic region 5     | 52518132-53027386               | 0.00001449   |
| Ampliconic region 6     | 55464117-55574172               | 0.00000691   |
| Ampliconic region 7     | 62335733-62495350               | 0.00000667   |
| Ampliconic region 8     | 70894117-71055682               | 0.00000457   |
| Ampliconic region 9     | 71941159-72325075               | 0.00001086   |
| Ampliconic region 10    | 100818723-100903977             | 0.00003235   |
| Ampliconic region 11    | 101435778-101774391             | 0.00002782   |
| Ampliconic region 12    | 103195105-103362341             | 0.00013457   |
| nonPAR- ampliconic regions | --                           | 0.00051647   |
| Low diversity region 1  | 10241177-12619185               | 0.00009100   |
| Low diversity region 2  | 16946047-18747389               | 0.00005870   |
| Low diversity region 3  | 19303480-22198160               | 0.00008547   |
| Low diversity region 4  | 38344992-41272675               | 0.00002217   |
| Low diversity region 5  | 45930478-77954462               | 0.00008178   |
| Low diversity region 6  | 99459295-111145964              | 0.00007904   |
| Low diversity region 7  | 128232540-136796526             | 0.00011957   |
| Low diversity region 8  | 151519514-155156362             | 0.00019179   |
| nonPAR- low diversity regions | --                           | 0.00060150   |
| nonPAR-ampliconic and low diversity regions | --                           | 0.00060150   |
Table S3. Complete Genomics unrelated genetic female samples. IDs, sex, population, ethnicity, and abbreviations are provided for each of the Complete Genomics samples used. In order to ensure there were no previously unreported relationships between individuals that might skew the analysis, we cross-checked each individual.  

| ID     | Sex   | Region      | Population  | Abbreviation |
|--------|-------|-------------|-------------|--------------|
| HG00732| Female| Americas    | PUERTO RICAN| PUR          |
| NA06985| Female| Europe      | UTAH/MORMON | CEU          |
| NA12004| Female| Europe      | UTAH/MORMON | CEU          |
| NA12890| Female| Europe      | UTAH/MORMON | CEU          |
| NA12892| Female| Europe      | UTAH/MORMON | CEU          |
| NA18502| Female| Africa      | YORUABA     | YRI          |
| NA18505| Female| Africa      | YORUABA     | YRI          |
| NA18508| Female| Africa      | YORUABA     | YRI          |
| NA18517| Female| Africa      | YORUABA     | YRI          |
| NA18526| Female| East Asia   | HAN CHINESE | CHB          |
| NA18537| Female| East Asia   | HAN CHINESE | CHB          |
| NA18555| Female| East Asia   | HAN CHINESE | CHB          |
| NA18942| Female| East Asia   | JAPANESE    | JPT          |
| NA18947| Female| East Asia   | JAPANESE    | JPT          |
| NA18956| Female| East Asia   | JAPANESE    | JPT          |
| NA19017| Female| Africa      | LUHYA       | LWK          |
| NA19129| Female| Africa      | YORUABA     | YRI          |
| NA19238| Female| Africa      | YORUABA     | YRI          |
| NA19648| Female| Americas    | MEXICAN ANCESTRY | MXL |
| NA19669| Female| Americas    | MEXICAN ANCESTRY | MXL |
| NA19701| Female| Africa      | AFRICAN ANCESTRY | ASW |
| NA19704| Female| Africa      | AFRICAN ANCESTRY | ASW |
| NA20502| Female| Europe      | TOSCANI (TUSCAN) | TSI |
| NA20847| Female| South Asia  | GUJARATI INDIAN | GIH |
| NA21733| Female| Africa      | MAASAI      | MKK          |
| NA21767| Female| Africa      | MAASAI      | MKK          |
**Table S4. Regions filtered from the X chromosome.** Regions of the human X chromosome that aligned with autosomal sequence in either chimpanzee or macaque were excluded from analysis. Regions of high divergence in the non-recombining regions (Supplementary Figure 2 and Supplementary Figure 3), were filtered out for supplementary analysis (Supplementary Table 2).

| Chr  | Start  | End    | Alignment | Reason for exclusion | Genes                                      |
|------|--------|--------|-----------|----------------------|--------------------------------------------|
| ChrX | 200815 | 388166 | Macaque   | Autosomal (Chr 7)    | PLCXD1(2), GTPBP6(1), LINC00685(2), PPP2R3B(1) |
|      |        |        |           |                      | VAMP7(5), IL9R(2), DDX11L16(1)            |
| ChrX | 155104073 | 155256767 | Macaque   | Autosomal (Chr 11, 13) |                           |
| ChrX | 155235958 | 155258790 | Chimpanzee | Autosomal (Chr 10, 15) | IL9R(2), DDX11L16(1)        |
| ChrX | 3000001 | 3300001 | Chimpanzee | High Divergence       | ARSF(3)                      |
| ChrX | 47900001 | 48300001 | Chimpanzee | High Divergence       | ZNF630(5), SSX(15), SPACAT(2) |
| ChrX | 48000001 | 48300001 | Macaque   | High Divergence       | SSX(14)                      |
| ChrX | 49100001 | 49400001 | Macaque   | High Divergence       | CCDC22(1), FOXP3(2), PPP1R3F(2), GAGE(68) |
| ChrX | 52200001 | 52900001 | Macaque   | High Divergence       | XAGE(33), SSX(14), SPANXN5(1)          |
| ChrX | 56700001 | 56900001 | Macaque   | High Divergence       | LINC01420(1), UQCRBP1(1)         |
| ChrX | 58000001 | 58100001 | Macaque   | High Divergence       |                           |
| ChrX | 64000001 | 64100001 | Macaque   | High Divergence       |                           |
| ChrX | 65500001 | 65800001 | Macaque   | High Divergence       |                           |
| ChrX | 70800001 | 71100001 | Macaque   | High Divergence       | BCYRN1(1), ACRC(1), CXCR3(2), CXorf49(4), LINC00891(1), LOC100132741(1) |
| ChrX | 75500001 | 75600001 | Macaque   | High Divergence       |                           |
| ChrX | 75800001 | 76100001 | Macaque   | High Divergence       | MIR325HG(7)                  |
| ChrX | 103100001 | 103400001 | Chimpanzee | High Divergence       | TMSB15B(1), H2BF(4), LOC(3), SLC25A53(1), ZCCHC18(2) |
| ChrX | 103200001 | 103400001 | Macaque   | High Divergence       | TMSB15B(1), H2BF(4), LOC(3), SLC25A53(1), ZCCHC18(2) |
| ChrX | 114900001 | 115100001 | Macaque   | High Divergence       | DANT2(2)                     |
| ChrX | 119100001 | 119200001 | Macaque   | High Divergence       | RHOF1(1)(2)                  |
| ChrX | 120000001 | 120100001 | Macaque   | High Divergence       | CT47(67)                     |
| ChrX | 134300001 | 134400001 | Macaque   | High Divergence       | CT55(2), ZNF75D(1)           |
| ChrX | 134700001 | 135000001 | Macaque   | High Divergence       | DDX26B(1), CT45(19), SAGE1(1) |
Table S5. Diversity across regions of the human X chromosome, with ampliconic and low diversity regions included. Diversity, measured as the average number of pairwise differences per site ($\pi$) between the X chromosomes of 26 unrelated genetic females, in each region of the human X chromosome is presented first unnormalized for mutation rate variation, then normalized using human-chimpanzee (hg19-panTro4) divergence and separately normalized for human-macaque (hg19-rheMac3) divergence. The regions analyzed include the pseudoautosomal region 1 (PAR1), pseudoautosomal region 2 (PAR2), X-transposed region (XTR) and the non-pseudoautosomal region either including the XTR (nonPAR) or excluding the XTR (nonPAR_{minus_XTR}). The ampliconic and low diversity regions have not been filtered out. $P$ values from permutation tests with 10,000 replicates are shown between each recombining and nonPAR region.

| Region          | Uncorrected $\pi$ | Divergence | $\pi$ Divergence | $\pi$ Divergence | $\pi$ Divergence | $\pi$ Divergence | $\pi$ Divergence |
|-----------------|-------------------|------------|------------------|------------------|------------------|------------------|------------------|
| nonPAR          | 0.000512          | 0.009814   | 0.053489         | 0.049702         | 0.010444         | 0.234423         | 0.002183         | 0.305070         | 0.001678         |
| nonPAR_{minus_XTR} | 0.000506          | 0.009782   | 0.053056         | 0.049512         | 0.011036         | 0.234372         | 0.002157         | 0.304460         | 0.001661         |
| PAR1            | 0.001505          | 0.022643   | 0.066482         | 0.099892         | 0.015070         | 0.337717         | 0.004457         | NA               | 0.000000         |
| $p$ vs. nonPAR  | **0.0000**        | 0.0388     | **0.0000**       | **0.0000**       | 0.0000           | **NA**           |                  |                  |                  |
| $p$ vs. nonPAR_{minus_XTR} | **0.0000**        | 0.0500     | 0.0000           | 0.0000           | 0.0000           | **NA**           |                  |                  |                  |
| PAR2            | 0.000643          | 0.008720   | 0.078348         | 0.040967         | 0.016677         | 0.218771         | 0.002940         | 0.257609         | 0.002497         |
| $p$ vs. nonPAR  | 0.1429            | 0.1342     | 0.1092           | 0.1298           | 0.1298           | 0.0947           |                  |                  |                  |
| $p$ vs. nonPAR_{minus_XTR} | 0.1553           | 0.1421     | 0.1132           | 0.1372           | 0.1372           | 0.0993           |                  |                  |                  |
| XTR             | 0.000747          | 0.010937   | 0.068256         | 0.056953         | 0.013108         | 0.245717         | 0.003038         | 0.336725         | 0.002217         |
| $p$ vs. nonPAR  | **0.0000**        | **0.0000** | **0.0000**       | **0.0000**       | **0.0000**       | **0.0000**       | **0.0000**       | **0.0000**       | **0.0000**       |
Supplemental References

1. R Core Team (2015). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).

2. Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F.A., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 14, 708–715.

3. Das, P.J., Chowdhary, B.P., and Raudsepp, T. (2009). Characterization of the bovine pseudoautosomal region and comparison with sheep, goat, and other mammalian pseudoautosomal regions. Cytogenet. Genome Res. 126, 139–147.

4. Gianfrancesco, F., Esposito, T., Montanini, L., Ciccodicola, A., Mumm, S., Mazzarella, R., Rao, E., Giglio, S., Rappold, G., and Forabosco, A. (1998). A Novel Pseudoautosomal Gene Encoding a Putative GTP-Binding Protein Resides in the Vicinity of the Xp/Yp Telomere. Hum. Mol. Genet. 7, 407–414.

5. (2002). Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc. Natl. Acad. Sci. U. S. A. 99, 16899–16903.

6. (2001). Protein phosphatase 2A interacts with the Src kinase substrate p130CAS. Publ. Online 21 Sept. 2001 Doi101038sjonc1204735 20,

7. Frederiksen, A.L., Hansen, S., Brixen, K., and Frost, M. (2014). Increased cortical area and thickness in the distal radius in subjects with SHOX-gene mutation. Bone 69, 23–29.

8. Zhang, W., Wang, J., Wang, Q., Chen, G., Zhang, J., Chen, T., Wan, T., Zhang, Y., and Cao, X. (2001). Identification of a novel type I cytokine receptor CRL2 preferentially expressed by human dendritic cells and activated monocytes. Biochem. Biophys. Res. Commun. 281, 878–883.

9. Vaz, C., Ahmad, H.M., Sharma, P., Gupta, R., Kumar, L., Kulshreshtha, R., and Bhattacharya, A. (2010). Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. BMC Genomics 11, 288.

10. Hara, T., and Miyajima, A. (1996). Function and Signal Transduction Mediated by the Interleukin 3 Receptor System in Hematopoiesis. STEM CELLS 14, 605–618.

11. Schiebel, K., Weiss, B., Wöhrle, D., and Rappold, G. (1993). A human pseudoautosomal gene, ADP/ATP translocase, escapes X-inactivation whereas a homologue on Xq is subject to X-inactivation. Nat. Genet. 3, 82–87.

12. Botros, H.G., Legrand, P., Pagan, C., Bondet, V., Weber, P., Ben-Abdallah, M., Lemière, N., Huguet, G., Bellalou, J., Maronde, E., et al. (2013). Crystal structure and
functional mapping of human ASMT, the last enzyme of the melatonin synthesis pathway. J. Pineal Res. 54, 46–57.

13. Fujiwara, S.-I., Yamashita, Y., Choi, Y.L., Watanabe, H., Kurashina, K., Soda, M., Enomoto, M., Hatanaka, H., Takada, S., Ozawa, K., et al. (2007). Transforming activity of purinergic receptor P2Y, G protein coupled, 8 revealed by retroviral expression screening. Leuk. Lymphoma 48, 978–986.

14. Jarnæss, E., Stokka, A.J., Kvissel, A.-K., Skålhegg, B.S., Torgersen, K.M., Scott, J.D., Carlson, C.R., and Taskén, K. (2009). Splicing Factor Arginine/Serine-rich 17A (SFRS17A) Is an A-kinase Anchoring Protein That Targets Protein Kinase A to Splicing Factor Compartments. J. Biol. Chem. 284, 35154–35164.

15. Gianfrancesco, F., Sanges, R., Esposito, T., Tempesta, S., Rao, E., Rappold, G., Archidiacono, N., Graves, J.A.M., Forabosco, A., and D’Urso, M. (2001). Differential Divergence of Three Human Pseudoautosomal Genes and Their Mouse Homologs: Implications for Sex Chromosome Evolution. Genome Res. 11, 2095–2100.

16. Esposito, T., Gianfrancesco, F., Ciccodicola, A., Montanini, L., Mumm, S., ’Urso, M.D., and Forabosco, A. (1999). A Novel Pseudoautosomal Human Gene Encodes A Putative Protein Similar to Ac-like Transposases. Hum. Mol. Genet. 8, 61–67.

17. Goodfellow, P., Pym, B., Mohandas, T., and Shapiro, L.J. (1984). The cell surface antigen locus, MIC2X, escapes X-inactivation. Am. J. Hum. Genet. 36, 777–782.

18. Ellis, N.A., Tippett, P., Petty, A., Reid, M., Weller, P.A., Ye, T.Z., German, J., Goodfellow, P.N., Thomas, S., and Banting, G. (1994). PBDX is the XG blood group gene. Nat. Genet. 8, 285–290.

19. Weller, P.A., Critcher, R., Goodfellow, P.N., German, J., and Ellis, N.A. (1995). The human Y chromosome homologue of XG: transcription of a naturally truncated gene. Hum. Mol. Genet. 4, 859–868.

20. Blanco-Arias, P., Sargent, C.A., and Affara, N.A. (2002). The human-specific Yp11.2/Xq21.3 homology block encodes a potentially functional testis-specific TGIF-like retroposon. Mamm. Genome Off. J. Int. Mamm. Genome Soc. 13, 463–468.

21. Blanco, P., Sargent, C.A., Boucher, C.A., Howell, G., Ross, M., and Affara, N.A. (2001). A Novel Poly(A)-Binding Protein Gene (PABPC5) Maps to an X-Specific Subinterval in the Xq21.3/Yp11.2 Homology Block of the Human Sex Chromosomes. Genomics 74, 1–11.

22. Yoshida, K., and Sugano, S. (1999). Identification of a Novel Protocadherin Gene (PCDH11) on the Human XY Homology Region in Xq21.3. Genomics 62, 540–543.
23. Watanabe, T.K., Fujiwara, T., Nakamura, Y., Hirai, Y., Maekawa, H., and Takahashi, E. (1996). Cloning, expression pattern and mapping to Xq of NAP1L3, a gene encoding a peptide homologous to human and yeast nucleosome assembly proteins. Cytogenet. Cell Genet. 74, 281–285.

24. Bonis, M.L.D., Cerase, A., Matarazzo, M.R., Ferraro, M., Strazzullo, M., Hansen, R.S., Chiurazzi, P., Neri, G., and D’Esposito, M. (2006). Maintenance of X- and Y-inactivation of the pseudoautosomal (PAR2) gene SPRY3 is independent from DNA methylation and associated to multiple layers of epigenetic modifications. Hum. Mol. Genet. 15, 1123–1132.

25. Vacca, M., Albania, L., Ragione, F.D., Carpi, A., Rossi, V., Strazzullo, M., De Franceschi, N., Rossetto, O., Filippini, F., and D’Esposito, M. (2011). Alternative splicing of the human gene SYBL1 modulates protein domain architecture of longin VAMP7/TI-VAMP, showing both non-SNARE and synaptobrevin-like isoforms. BMC Mol. Biol. 12, 26.

26. Chang, M.S., Engel, G., Benedict, C., Basu, R., and McNinch, J. (1994). Isolation and characterization of the human interleukin-9 receptor gene. Blood 83, 3199–3205.

27. Costa, V., Casamassimi, A., Roberto, R., Gianfrancesco, F., Matarazzo, M.R., D’Urso, M., D’Esposito, M., Rocchi, M., and Ciccodicola, A. (2009). DDX11L: a novel transcript family emerging from human subtelomeric regions. BMC Genomics 10, 250.

28. Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2015). The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 43, D670–D681.

29. Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., et al. (2010). Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. Science 327, 78–81.