Supplementary Note

I. Pseudogenes on the human or chimpanzee Y chromosomes misclassified as genes by Kuroki *et al.* [1]
II. Sequencing error rates

I. Pseudogenes on the human or chimpanzee Y chromosomes misclassified as genes by Kuroki *et al.* [1] Apart from these cases, the lists of human and chimpanzee Y-linked genes published by Kuroki *et al*. are in agreement with prior reports[2,3].

Chimpanzee *TMSB4Y*

Located on both the human and chimpanzee Y chromosomes, *TMSB4Y* was classified in prior reports as an intact, actively transcribed gene on the human Y chromosome[2,4] but as an inactivated pseudogene on the chimpanzee Y chromosome[3]. While human *TMSB4Y* is clearly transcribed[2,4], we tested for but found no evidence of transcription of *TMSB4Y* in chimpanzee[3]. Moreover, in chimpanzee, the single splice donor site within the *TMSB4Y* coding region has been lost through mutation (see Supplementary Fig. 8 in ref. 3.) Kuroki *et al.*[1] classified *TMSB4Y* as an intact, functional gene not only in human but also in chimpanzee.

Chimpanzee *USP9Y*

Located on both the human and chimpanzee Y chromosomes, *USP9Y* was classified in prior reports as an intact, actively transcribed gene on the human Y chromosome[2,4] but as an inactivated pseudogene on the chimpanzee Y chromosome[3]. In the chimpanzee Y chromosome, several splice sites in *USP9Y* have been lost through mutation and, as confirmed by sequencing of chimpanzee cDNAs, its open reading frame (ORF) is severely disrupted due to exon skipping[3]. While human *USP9Y* has an ORF of 7664 bp, the longest ORF in the chimpanzee gene would encode a protein of only 675 amino acids, with no intact catalytic domain[3]. Kuroki *et al*.[1] classified *USP9Y* as an intact, functional gene not only in human but also in chimpanzee.

Human and chimpanzee *GYG2*

Located on both the human and chimpanzee Y chromosomes, *GYG2* was classified as a pseudogene both in the prior report of the sequence of the human Y chromosome[2] and in the prior report of the sequence of the X-degenerate region of the chimpanzee Y chromosome[3]. Like most genes and many pseudogenes in the X-degenerate region, *GYG2* has a homolog on the human X chromosome. This X-linked homolog has 12 exons and an ORF of 1505 bp. Comparison with this X-linked homolog reveals that *GYG2* is grossly disrupted on both the chimpanzee and human Y chromosomes: *GYG2* lacks exons 5 through 12 and its ORF is truncated to 242 bp. Ten ESTs in GenBank match the human Y-linked pseudogene, but all but one of these are from cancers. We sought but failed to detect transcription of *GYG2* in normal human tissues[2]. Kuroki *et al.*[1]

provided no evidence that *GYG2* is transcribed in normal human or chimpanzee tissues, but classified it as an intact, functional gene in both species.

Human *CD24L4*

Located on the human Y chromosome, *CD24L4* is a processed (intronless) derivative of an actively transcribed autosomal gene, *CD24*. *CD24L4* is the result of a retrotransposition event that occurred after the divergence of the human and chimpanzee lineages. We are aware of no bona fide evidence that *CD24L4* is transcribed. In the absence of such evidence, it is likely that *CD24L4* is a processed pseudogene. Kuroki *et al.*[1] cited no evidence that *CD24L4* is transcribed but classified it as a functional gene.

Confusion with respect to the transcriptional status of *CD24L4* may stem from the fact that the nucleotide divergence between *CD24L4* and its autosomal progenitor, *CD24*, is only 1.14%. GenBank contains five cDNA entries whose nucleotide sequences perfectly match the *CD24L4* genomic sequence.

Invitrogen cDNA clone number	*CD24L4*-matching accession number	*CD24*-matching accession numbers	Accession numbers for clone sequences we obtained
CS0DF007YJ17	CR623460	AL535013, AL535014	DQ530230
CS0DF014YG04	CR609933	AL536277, AL566271	DQ530231
CS0DF027YP12	CR608613	AL538376, AL538375	DQ530233
CS0DI064YI12	CR598378	AL551360, AL574949	DQ530234
CS0DF027YF12	CR591421	AL538366, AL538367	DQ530232

These deposited cDNA sequences, however, are erroneous. The deposited sequences are annotated as deriving from five different Invitrogen cDNA clones. Each of the five cDNA clones is also the source for two other GenBank cDNA deposits that are nearly identical to the *CD24* (autosomal) genomic sequence and quite distinct from the *CD24L4* (Y-chromosomal) genomic sequence. At the Invitrogen website (http://clones.invitrogen.com) , the *CD24*-matching accession numbers are the sole sequences attributed to the clones; there are no records linking *CD24L4* to these cDNA clones. Suspecting that the *CD24L4*-matching sequences were erroneous deposits to GenBank, we ordered the five cDNA clones from Invitrogen and sequenced them in our laboratory. All five clones are indeed derived from *CD24* and not *CD24L4* (Supplementary Figure). With respect to EST evidence, there are only two ESTs in GenBank that perfectly match *CD24L4* (accession numbers BG428156 and DB148174). By contrast, there are 2137 EST matches for *CD24* listed in UniGene. We do not believe that two EST sequences provide adequate support to the claim that *CD24L4* is a gene.

II. Sequencing error rates

Error rate from Hughes, *et al.*[3]: 1/204,000 nucleotides

Error rate from Kuroki, *et al.*[1]: 1/500,000 nucleotides

Combined error rate = 1/204,000 + 1/500,000 = 1/145,000 nucleotides

Thus errors in one of the two sequencing studies may account for about one in every three substitutions that appear to differentiate the Y chromosomes of PTB1 and CHORI-251 (estimated divergence of 1/50,000 nucleotides).

1. Kuroki, Y. et al. Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway. *Nat Genet* (2006).
2. Skaletsky, H. et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825-37 (2003).
3. Hughes, J. F. et al. Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* **437**, 100-3 (2005).
4. Lahn, B. T. & Page, D. C. Functional coherence of the human Y chromosome. *Science* **278**, 675-80 (1997).