List of supplemental materials

Sequencing human-gibbon breakpoints of synteny reveals mosaic new insertions at rearrangement sites

Santhosh Girirajan[1*], Lin Chen[1*], Tina Graves[2], Tomas Marques-Bonet[1], Mario Ventura[3], Catrina Fronick[2], Lucinda Fulton[2], Mariano Rocchi[3], Robert S. Fulton[2], Richard K. Wilson[2], Elaine R. Mardis[2] and Evan E. Eichler[1†]

1. **Supplemental Figures 1–7**

2. **Supplemental File 1**: Sequence resolution of human-gibbon breakpoints. Class 1 and class 2 breakpoints are shown. The site-specific recombinogenic and double strand break-associated sequences are also highlighted: known recombination hotspot 5-9mers (*shaded blue*) {Myers, 2005 #57}, topoIIv and topoIId sites (*black box*) {Sander, 1985 #82; Spitzner, 1988 #80}, immunoglobulin class switch repeats (*blue underline*) {Abeysinghe, 2003 #9}, and translin binding site (*black underline*) {Aoki, 1995 #100; Kanoe, 1999 #83}. The block arrows in class 2 breakpoints indicate insertions (LINE-green, LTR-orange, SINE-purple, segmental duplication-gray, simple repeats-black (AT-rich elements).

3. **Supplemental File 2**: Nucleotide variations in human gene orthologs *CSH1*, *CSH2* and *CSHL*.

4. **Supplemental File 3**: Phylogenetic trees for all the genes at or in the vicinity of the breakpoints. d_N/d_S (ω) values, synonymous and non-synonymous substitutions are also shown.

5. **Supplemental File 4**: Analysis of gibbon clones reported by Carbone et al, 2006. Modified miropeat images show the orthologous human chromosomal regions mapping to gibbon BACs containing synteny breaks.

6. **Supplementary data**: This document contains Miropeat analysis of all the 24 human gibbon breakpoints.

7. **Supplementary Tables** 1 - 11

Legend to Supplementary Figures

Supplementary Figure 1. LTR insertions at human-gibbon synteny breakpoints. (a) Human-gibbon pairwise alignment by *Miropeat* analysis is shown. The breakpoint interval is expanded (*dashed lines*) to show the insertion of the LTR and incomplete LINE elements. Note the retroviral integration target site duplication (*blue box*) flanking the insertion site {Jin, 2002 #98; Yohn, 2005 #97}. The orange letters denote LTR and green letters indicate LINE-1 elements. **(b)** A three-way sequence alignment of a human-gibbon synteny breakpoint containing the LTR insertion is shown. Also note the flanking target site duplication (*blue box*).

Supplementary Figure 2: Mosaic insertions at the breakpoints. The schematic shows human-NLE gibbon pairwise alignment with the breakpoint insertion containing a "hodgepodge" or a mosaic of SINE (*purple*), LTR (*orange*), and LINE (*green*) indicated by a stick diagram. Note also the presence of a 16-mer "filler" sequence.

Supplementary Figure 3. Decreased frequency of gene disruption at the human-gibbon synteny breaks. The number of gene disruptions observed in a simulation of 50 breaks in the human genome (hg17/Build 35) simulated at least 100 times is significantly (*P*=0.02) more than the expected (n=7) number of breaks. Consistent with our initial bias, we excluded segmental duplications from the human genome sequences used for simulation.

Supplementary Figure 4. Segmental duplication-mediated events in the vicinity of human-gibbon synteny breaks. Alignments between NLE BAC sequences and human chromosomes show three examples of SD-mediated events. (**a**) Note the insertion of the *DZIP* gene, contained within the segmental duplication, into the gibbon BAC. Several exons of the gene, including the start codon, are missing rendering this a pseudogene. (**b**) Vomeronasal receptor genes, *VN1R2* and *VN1R4*, are shown in relation to other genes in the gibbon BAC. These genes are inactive in chimpanzee, gorilla, and in gibbons. Also shown is the *DEPDC4* gene, disrupted at the break.

Supplementary Figure 5. Paralogous genes in the vicinity (<50 kbp) of human-gibbon break points. Sequence alignments show regions of synteny as well as the breakpoints in gibbon BAC compared to humans. Paralogous gene clusters for (**a**) growth hormone (*GH2*) and (**b**) acyl Co-A thioesterase genes (*ACOT*), mapping within segmental duplication, are shown. The segmental duplications in both cases map on one side of the breakpoint. Also shown are genes (arrows depicting transcriptional orientation) and sequence repeat tracks. For the repeat tracks, LINE (*green*), SINEs (*purple*), LTRs (*orange*), and others (*black*). Both the segmental duplications (*red arrow*) shown are shared between humans and gibbons.

Supplementary Figure 6. Microhomology-mediated nonhomologous recombination mechanism. The schematic shows two human chromosomal strands containing microhomologous sequences. A 5′-3′ exonuclease trims the ends of the DNA to expose the microhomology strands. Annealing of the two strands creates a new “rearranged” chromosome.

Supplementary Figure 7. Mosaic insertions at human gibbon synteny breaks. A summary of all the mosaic insertions observed in our analysis is shown. The different repetitive segments are color coded as LINE (green), SINE (*purple*), LTR (orange), and disparate sequences (*red, pink, dark green, light blue*) forming SD (4.3 kb and 20 kb mosaics) are also represented.