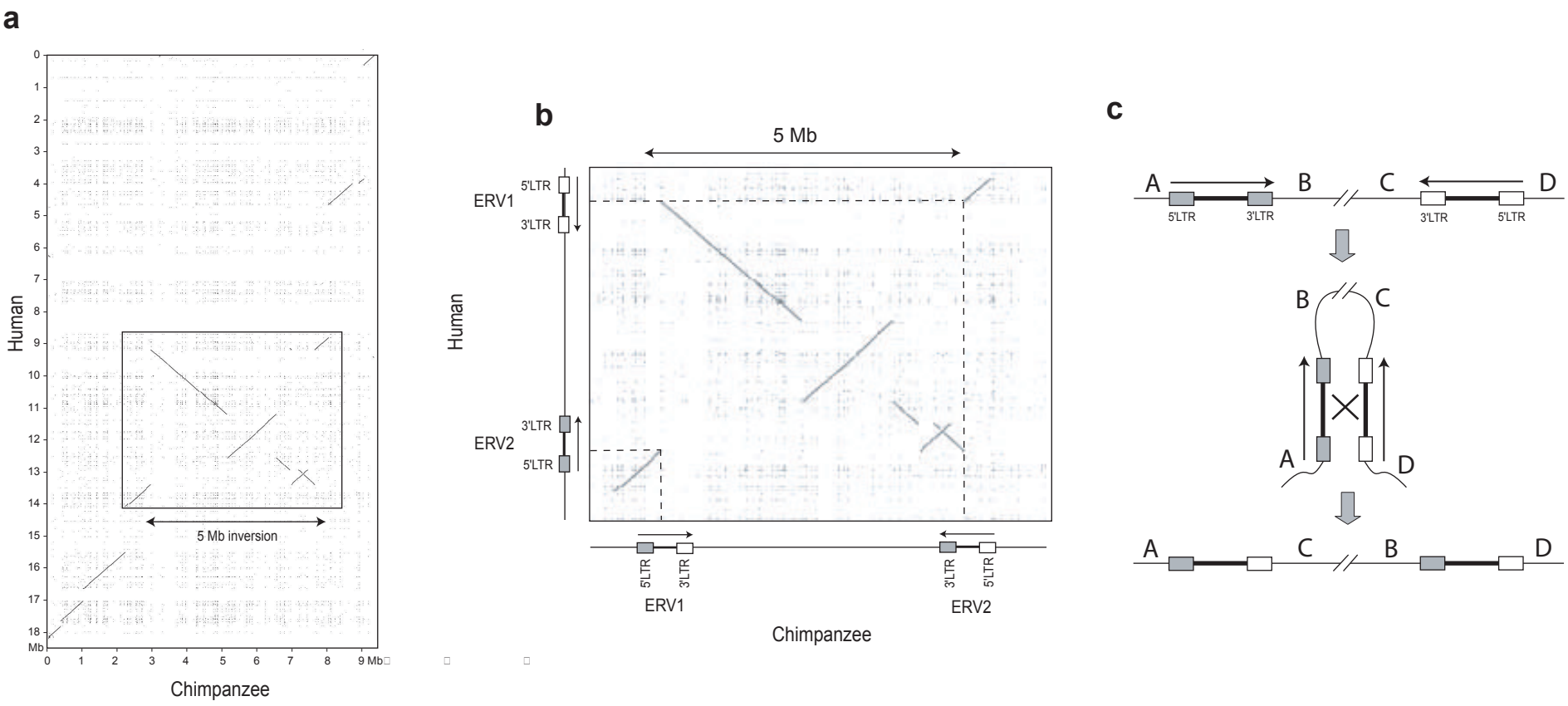
Supplementary Figure 2



Supplementary Figure 2 5-Mb inversion in the chimpanzee lineage mediated by homologous recombination between endogenous retroviruses (ERVs). **a.** Dot-plot from Supplementary Figure 1. The boundaries of the 5-Mb inversion are boxed. **b.** The area in the box from part a is enlarged. The dotted lines show the positions of the inversion breakpoints on the axes. The relative positions of the ERV elements in the chimpanzee and human sequences are shown next to each axis (not to scale). The LTRs of the ERVs are represented as boxes. The orientation of each ERV is indicated by an arrow. **c.** Schematic model of the recombination event. Letters A-D denote the Y chromosomal sequences located upstream and downstream of each element. **d.** Neighbor-joining tree of chimpanzee and human ERV1 and ERV2 LTR sequences. Sequences were aligned using the CLUSTAL W algorithm in MacVector 7.2.2 and corrected by hand, and PAUP 4.0*b10 was used for the phylogenetic analysis. **e.** Schematic representations of the two ERV elements in each species and its flanking 4 bp repeats.

There are two full-length endogenous retroviruses (referred to as ERV1 and ERV2) in inverted orientation at the breakpoints in both species (b), indicating that both integration events occurred in a common ancestor of humans and chimpanzees. A recombination event between these two inverted repeat elements would cause an inversion of the intervening sequence, which in this case spans over 5 Mb (c). The ERV elements in the chimpanzee sequence, but not the human, bear the hallmarks of recombination. Recombinant ERVs can be readily identified in two ways: long terminal repeat (LTR) divergence and mismatched target site duplications[1]. First, ERVs have LTRs at their 5' and 3' termini. These sequences are usually a few hundred base pairs in length and are identical at the time of viral integration. After incorporation into the genome, the ERV acquires mutations over time, and the LTR sequences diverge. However, given the rarity of such integration and fixation events, the LTRs of a single ERV should be more similar to each other than they are to the LTRs of any other ERV element in the genome. In a neighbor-joining phylogenetic analysis of the LTR sequences from both ERVs in both species, the deepest node in the tree separates ERV1 and ERV2 and the subsequent nodes separate the LTRs within each ERV (d). While the two flanking LTRs (referred to as 5' and 3') of both human elements cluster together, the chimpanzee LTRs cluster with the opposite element. The 5' LTR of ERV2 and the 3' LTR of ERV1 cluster together, while the 5'LTR of ERV1 and the 3'LTR of ERV2 cluster together. The topology of this tree is fully supported by 100 bootstrap replicates. Another hallmark of a recombined retrovirus is mismatched flanks. The process of retroviral integration generates a short (4-5 bp) duplication of the target sequence. While both target site duplications for the human ERVs match, the chimpanzee's do not and are in fact mismatched, supporting the phylogenetic evidence for the rearrangement (e).

1. Hughes, J. F. & Coffin, J. M. Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nat Genet* 29, 487-489 (2001).