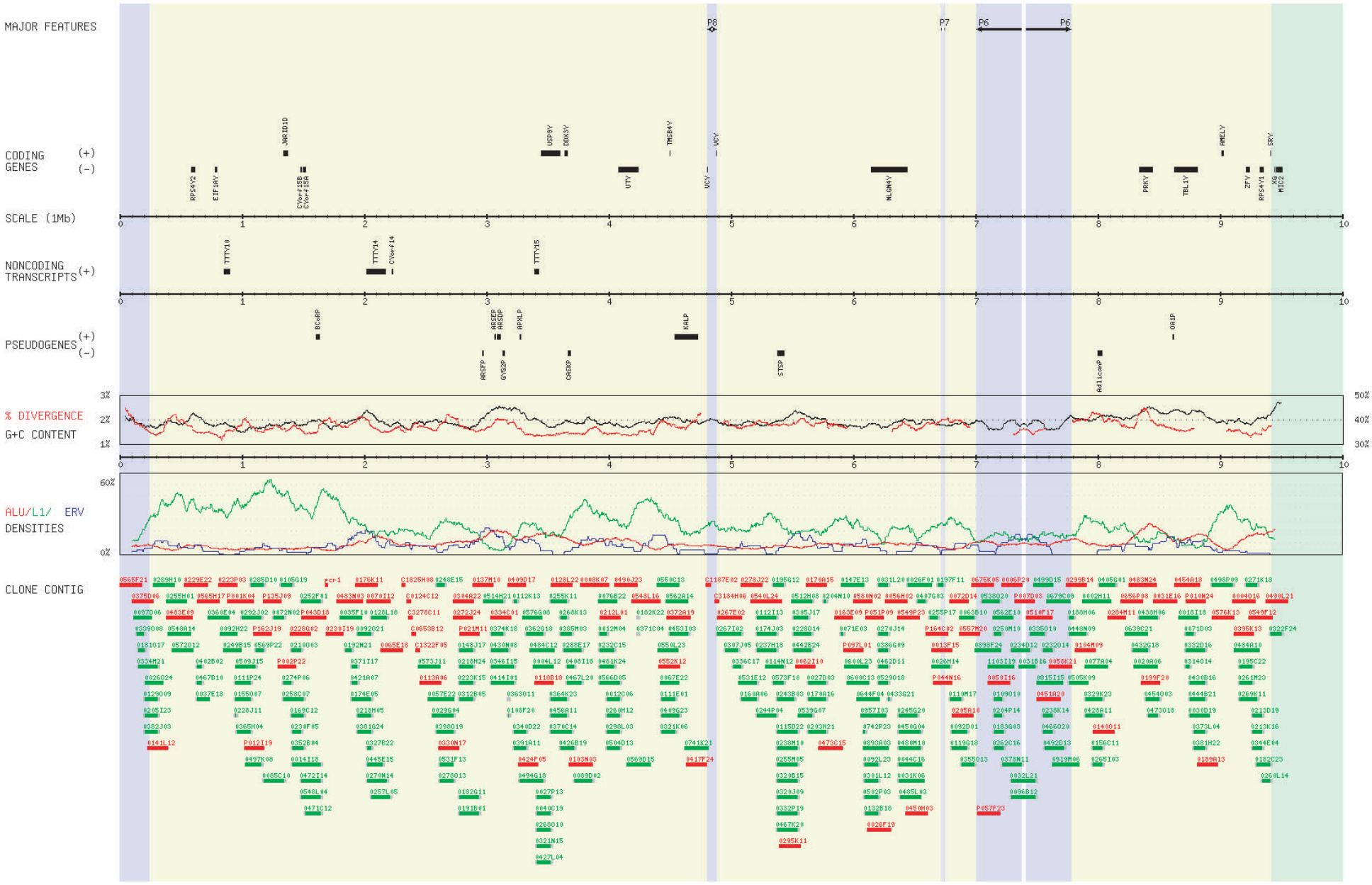
Supplementary Figure 5



Supplementary Figure 5 Sequence-based map of the chimpanzee X-degenerate sequence. This map provides a detailed view of the chimpanzee sequence shown in the dot-plot figures (Figure 1 and Supplementary Figure 1). The background colors represent the three different sequence classes present in this region: X-degenerate (yellow), ampliconic (blue), and pseudoautosomal (green). All sequence features and BACs are drawn to scale. a. Palindromes 6, 7 and 8 (P6-8) are shown. Paired arrows represent the arms of the palindromes, and the space between paired arrows represents spacer regions present in these structures. b. The positions of sixteen coding genes that are actively transcribed in humans are shown. Plus (+) strand above, minus (-) strand below. c. The positions of non-coding transcription units are shown. Transcription of these sequences has been verified by RT-PCR analysis, but there is no evidence for coding potential of these transcripts. d. The positions of 11 pseudogenes are shown. Plus (+) strand above, minus (-) strand below. e. Percent divergence (red) with the corresponding human sequence was calculated in a 100 kb sliding window with 1 kb steps. Gaps in the plot represent regions in the chimpanzee sequence that are deleted in the human Y chromosome. Percent G+C content (black) of the chimpanzee sequence was calculated using the same sliding window and step size. The Y chromosome divergence level is positively correlated with GC content along the length of the sequence ($r=0.28$, $P<0.0001$). f. Repeats were identified using RepeatMasker and the total percentage of the sequence composed of Alu (red), L1 (green), and ERV (blue) elements was calculated in a 200 kb sliding window with 1 kb steps. g. Chimpanzee Y chromosome BAC and fosmid clones identified in analysis. Labels are library clone names. Names with no alphabetic prefix are from CHORI-251. The prefix "P" designates BACs from the RPCI-43 library. The prefix "C" designates fosmids from the CHORI-1251 library. PCR (designated "pcr1") was used to confirm the sequence of two neighboring BACs that abut precisely with no overlap. The BACs and fosmids shown in red are the set of 90 overlapping clones that were sequenced completely. The two gaps that remain in the sequence are between fosmids C1825H08 and C0124C12 at approximately 2.3 Mb, and between fosmid C1187E02 and fosmid C3184H06 at approximately 4.8 Mb. The green BACs are those that were mapped to the Y using STS content mapping. In some cases, the BAC end sequences are known, so their positions on the Y was determined precisely. Grey shading at one or both ends of a BAC indicates that the end was not sequenced, so its position is approximated by STS content.