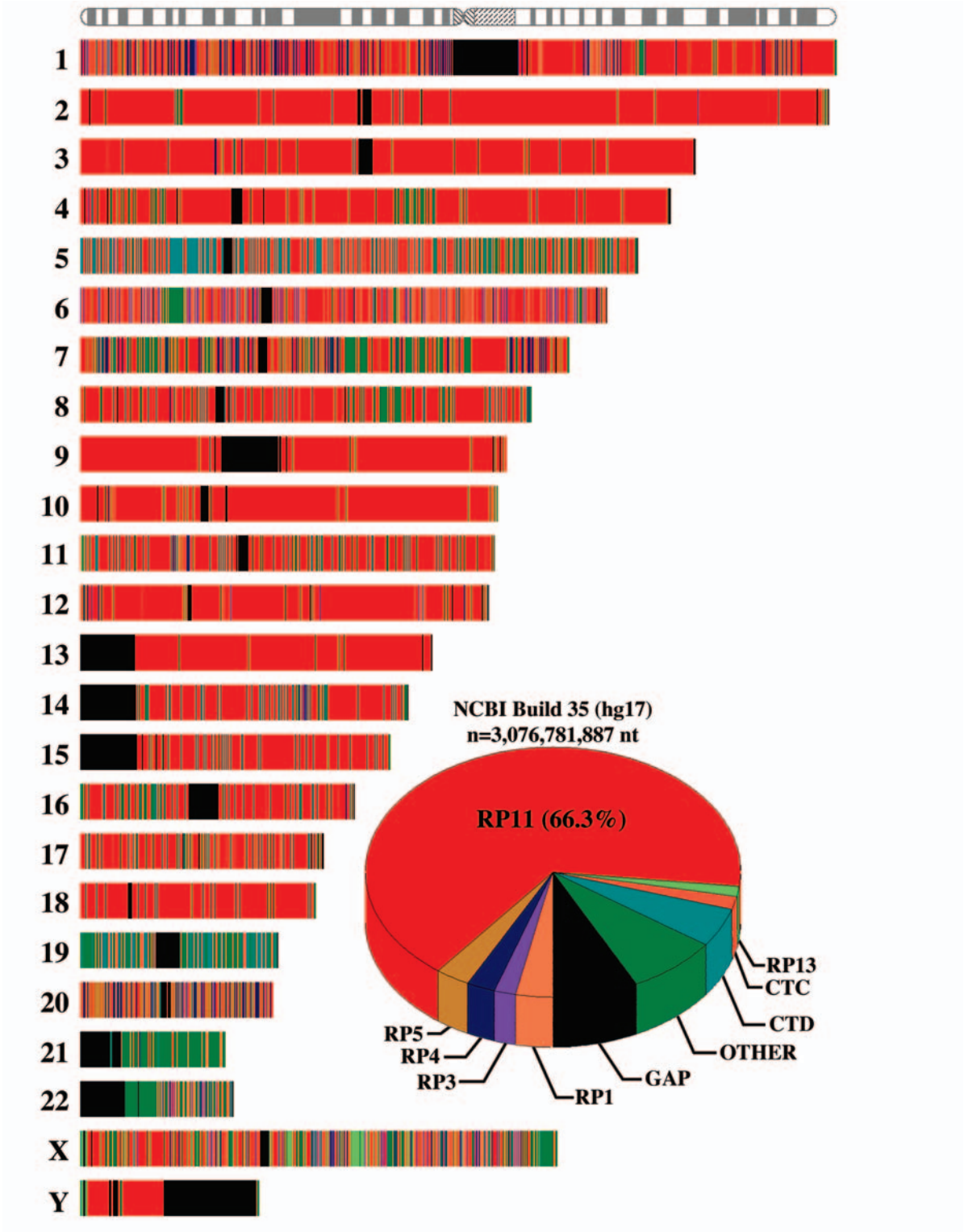
In format provided by Feuk *et al.* (February 2006)



S1 | **Composition of the human reference assembly.** The most recent assembly of the human genome sequence, which is used by annotation browsers such as NCBI (www.ncbi.nlm.nih.gov), UCSC (www.genome.ucsc.edu), and Ensembl (www.ensembl.org) is named hg17 (Build 35). This 'reference' assembly has a content of 3,076,781,887 nucleotides and is mainly (84.9%) comprised of eight genomic libraries from diploid DNA sources (libraries are RP13, RP11, RP5, RP4, RP3, RP1, CTC, CTD). Thirty-four other genomic libraries, along with 706 non-standard genome-sequencing library sources (for example, phage, cosmid and YAC-derived clones), comprise another 7.8% of the assembly. Information on these clone libraries is available at www.ncbi.nlm.nih.gov/genome/clone/. The composition of each chromosome in the assembly, based on DNA source, is shown with colours matching the pie-chart. An estimated 7.3% of sequence at 341 sites (including centromeres, acrocentric regions, and cloning/sequencing gaps) is thought to be missing from this assembly, as indicated in black. Structural variants that coincide with the boundaries of sequences from different contributing chromosomes will complicate interpretation, as will chromosomal loci that are not represented in the sequence assembly.