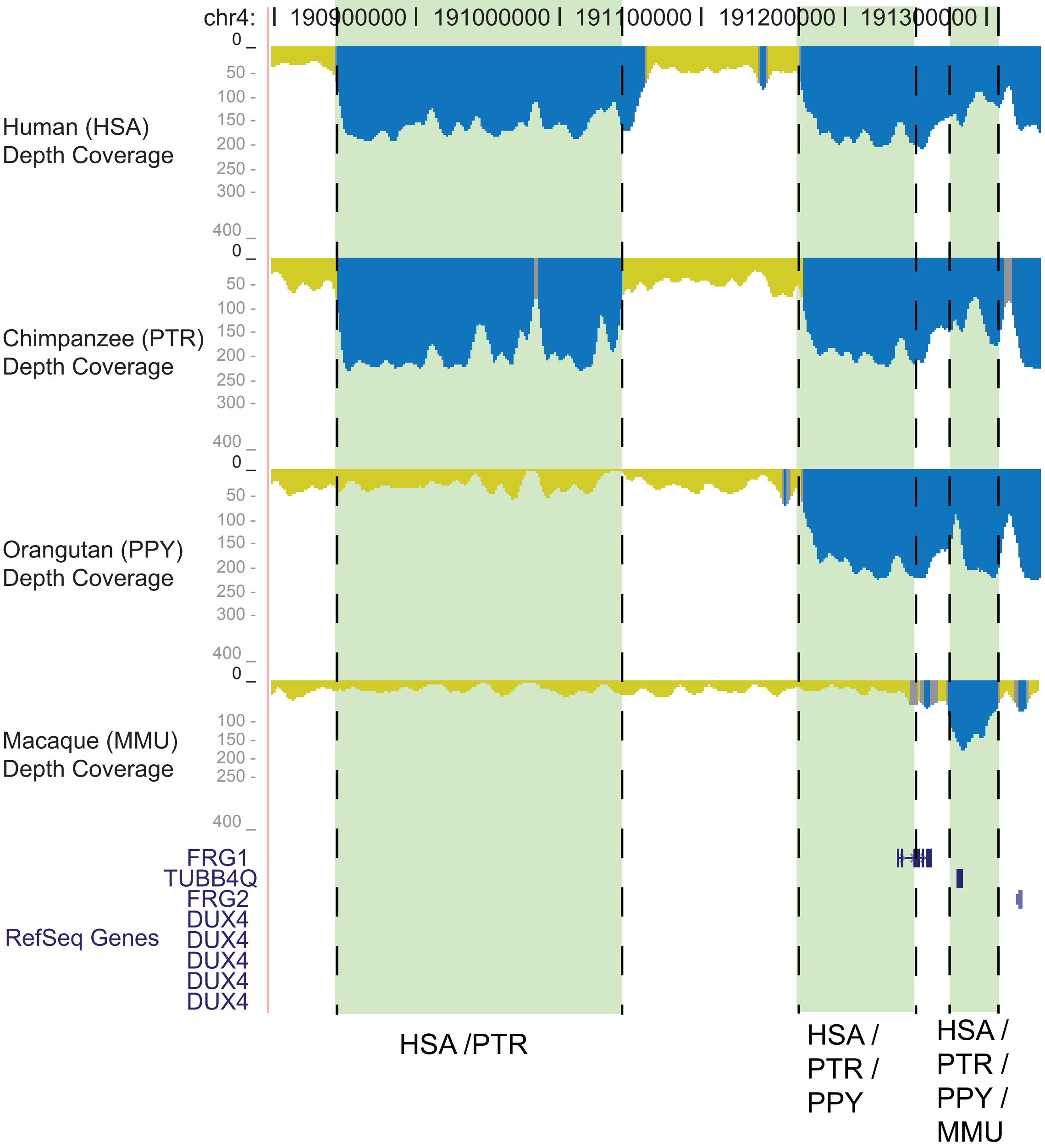


Fig. S1: Comparative primate segmental duplication analysis. The figure shows how segmental duplications were classified based on the WSSD computational analysis within the context of the UCSC genome browser. A ~500 kbp region is depicted corresponding to the fascioscapulohumeral muscular dystrophy on human chromosome 4. The depth of sequence read coverage (number of reads in five kbp windows is shown for (human (HSA), chimpanzee (PTR), orangutan (PPY) and macaque (MMU)) based on the alignment of these reads against the human genome. Regions of excess depth of coverage (blue, putative duplication) contrast with regions showing a depth-of-coverage within 3 s.d. of the mean of single copy regions (yellow). Three examples of SD classification are shown: from left to right, a HSA/PTR shared duplication, a HSA/PTR/PPY shared duplication (including partially the FRG1 gene) and a HSA/PTR/PPY/MMU SDs (including the full TUBB4Q gene). Any duplicated primate region can be viewed along with supporting experimental data using our customized map of primate segmental duplications displayed on a UCSC browser mirror (http://humanparalogy.gs.washington.edu).

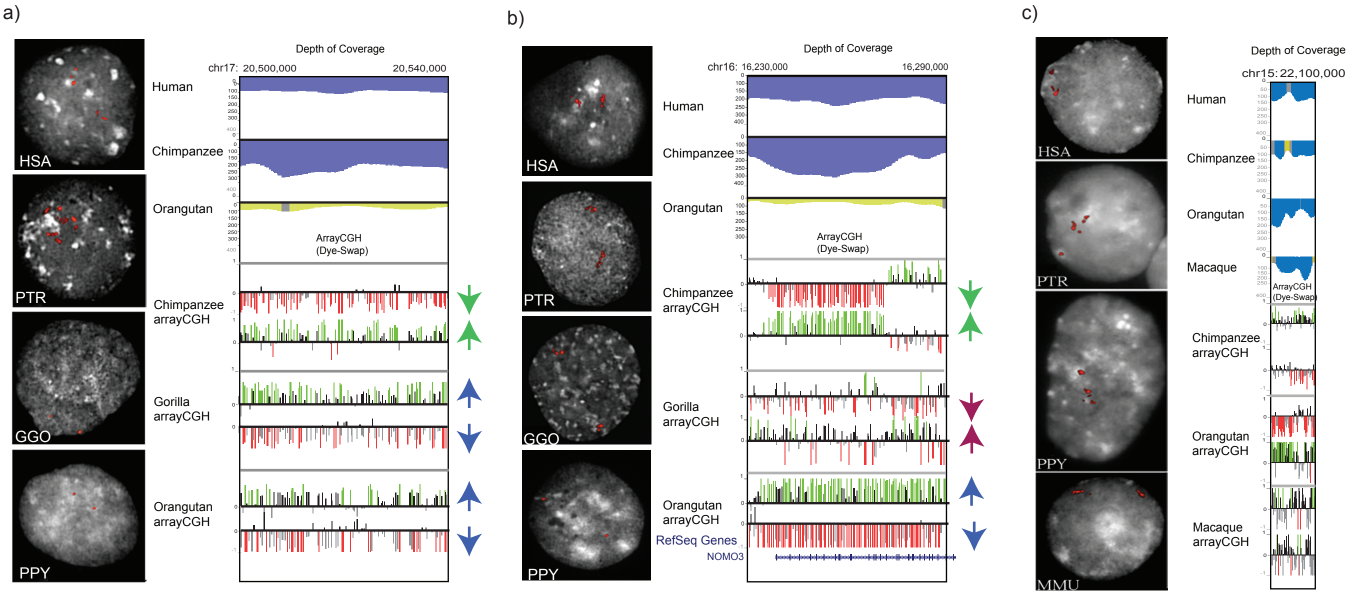


Fig. S2: FISH vs. cross-species arrayCGH data. This figure shows the specificity of our combined computational and experimental approach. **a)** An example of a human-chimpanzee shared duplication (predicted by WSSD analysis) that is single copy in gorilla and orangutan as determined by FISH and arrayCGH data (a replicate experiment is shown for each arrayCGH experiment as a dye-swap with the human reference DNA sample and test non-human primate DNA sample). **b)** An arrayCGH experiment showing a human-chimpanzee shared duplication that is duplicated in gorilla but single copy in orangutan based on experimental validation. **c)** A complex region mapping to one of the breakpoints of the Prader-Willi Syndrome that is duplicated in all four primate species showing patches of shared and species-specific duplications. ArrayCGH and FISH results confirm copy-number differences in regions of shared duplication (the region shown correspond to the extent of fosmid probe used in FISH experiment (WIBR2-0877G19)).

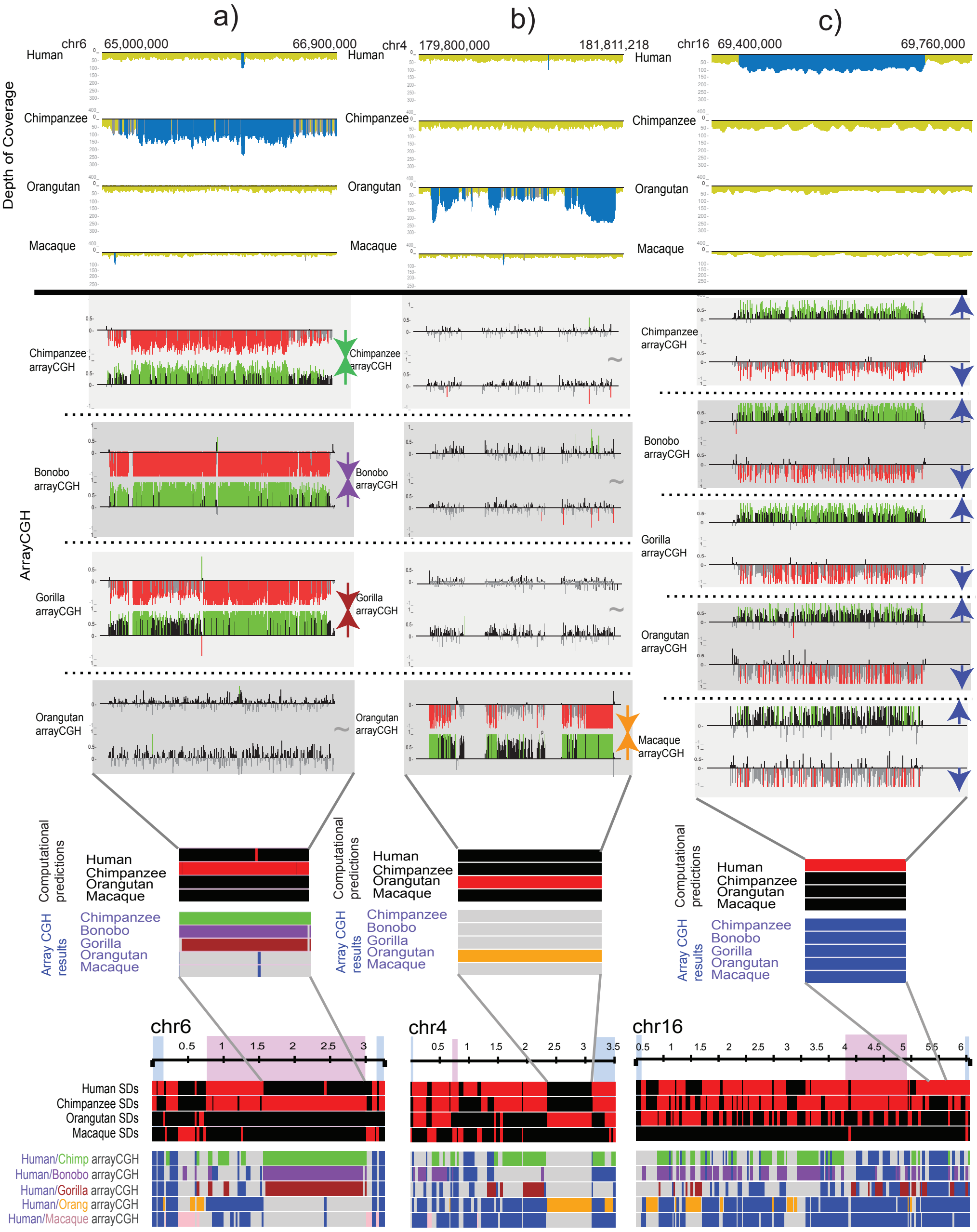


Fig. S3: Construction of a primate segmental duplication map. We combined computational and experimental predictions to construct a primate segmental duplication map on the human reference genome. Three real examples are shown depicting **a)** a chimpanzee-specific duplication, **b)** an orangutan-specific duplication and **c)** a human-specific duplication. The top panel shows the “in silico” prediction (WSSD computational analysis) while the middle panel shows the results by replicate dye-swap arrayCGH for each non-human primate against human. These results were concatenated across the genome and summarized in the duplication map (Fig. 2) as follows: Regions of segmental duplication are shown in red while black denotes single copy sequence in each of the species. The next 5 rows summarize the results of cross-species arrayCGH hybridization experiments. Regions of increased signal intensity in human (blue) contrast with regions of increased signal intensity in each of the nonhuman primate species: green (chimp), purple (bonobo), dark red (gorilla), orange (orang) and pink (macaque). Grey regions show no significant difference in signal intensity. The extent of pericentromeric duplications (<5 Mb of centromere) and subtelomeric (<1000 kbp) are highlighted in purple and blue respectively based on human genome organization.

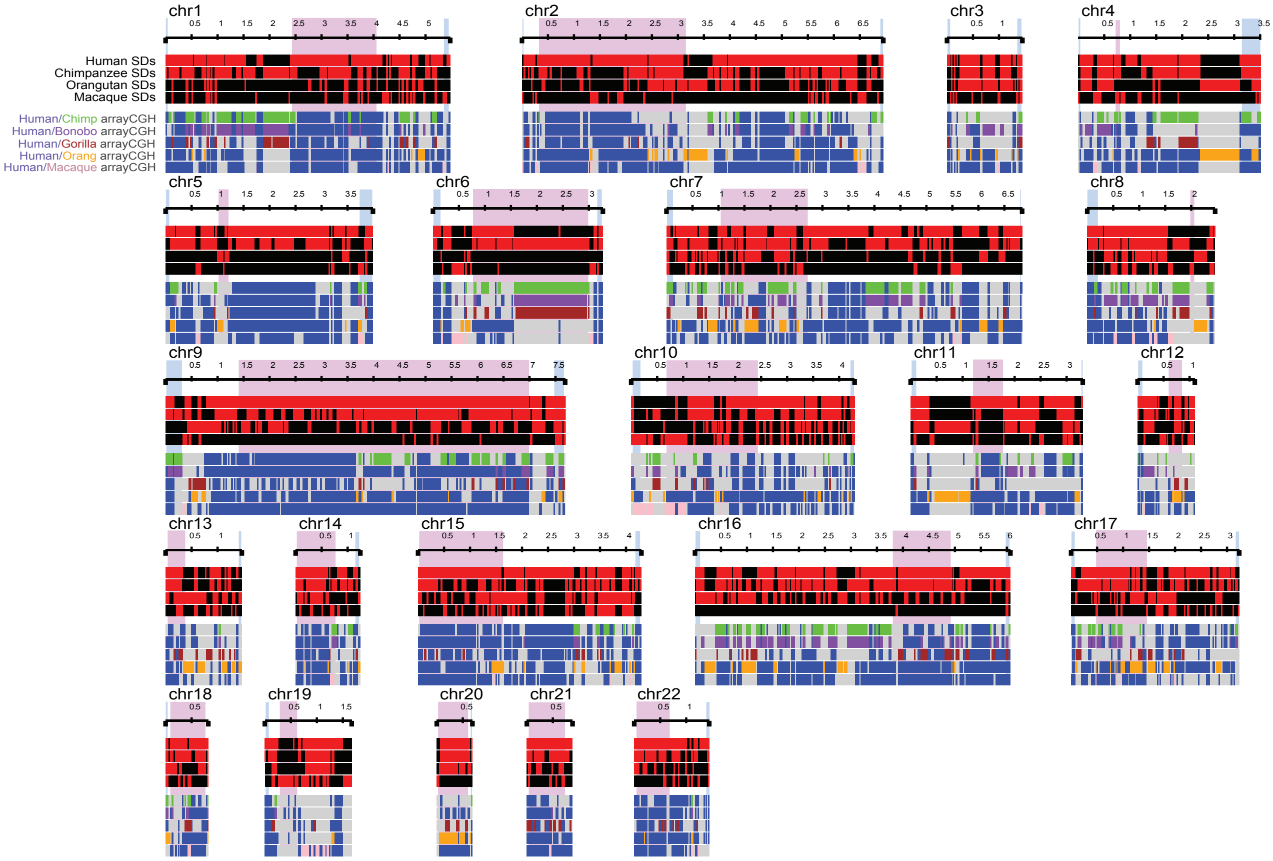


Fig. S4: Comparative primate duplication map. Computationally predicted regions of SDs (>20 kbp) (human, HSA; chimpanzee, PTR; orangutan, PPY and macaque, MMU) were concatenated and compared based on the human reference sequence (build35). SDs are shown in red while black denotes single copy sequence. The next 5 rows summarize the results of cross-species arrayCGH hybridization. Regions of increased signal intensity in human (blue) contrast with regions of increased signal intensity in chimp (green), bonobo (purple), gorilla (dark red), orang (orange) and pink (macaque). Grey regions show no significant difference in signal intensity (Fig. S3 for a schematic representation of the construction of the duplication map). Pericentromeric duplications (<5 Mb of centromere) and subtelomeric (<1000 kbp) are highlighted in purple and blue respectively based on human genome organization.

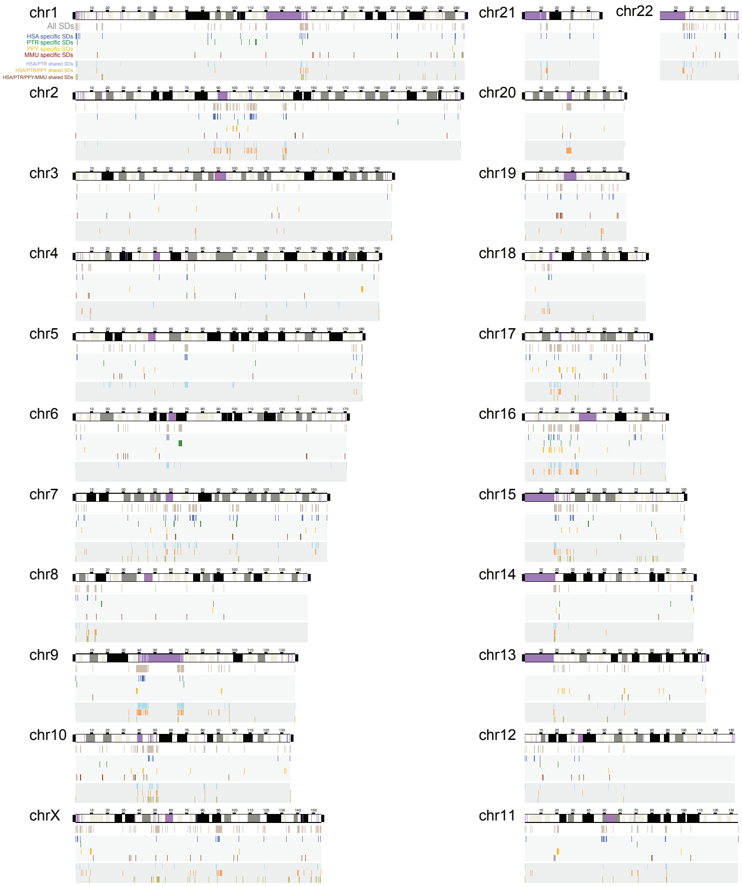
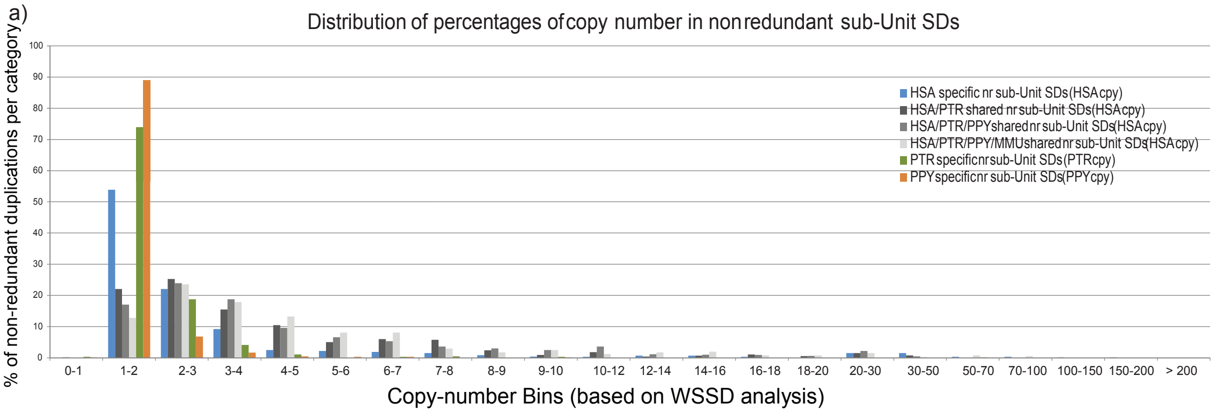


Fig. S5: Landscape of great-ape and human SDs in the human genome. The map shows the actual distribution of all great-ape SDs (>20 kbps) placed in the context of the human genome (build35). For each human chromosomal ideogram, there are 8 rows grouped by grey blocks into 3 groups: **a)** The union of all SDs; **b)** Species-specific SDs, from 2nd to 5th row, human (HSA), chimpanzee (PTR), orangutan (PPY) and macaque (MMU) specific duplications respectively; and **c)** Shared SDs, from 6th to 8th row, HSA/PTR, HSA/PTR/PPY and HSA/PTR/PPY/MMU duplications. Duplications cluster within the pericentromeric and subtelomeric regions as well as other regions of the genome.



b)

Category APE SDs	Total #N	Sum of sub-unit Size	Average Size of sub-Unit SDs	Median of Cpy_HSA	Median of Cpy_PTR	Median of Cpy_PPY
HSA specific sub-Unit SDs	740	7,268,951	9,822.90	1.94***	1.12	0.88
HSA/PTR shared sub-Unit SDs	784	8,247,488	10,519.75	3.11	2.77	0.93
HSA/PTR/PPY shared sub-Unit SDs	698	3,608,545	5,169.83	3.41	3.19	2.86
HSA/PTR/PPY/MMU shared sub-Unit SDs	408	1,815,531	4,449.83	3.70	3.51	2.81
PTR specific SDs	1,086	11,181,508	10,296.04	1.07	1.78	1.09
PPY specific SDs	3,633	30,348,197	8,353.48	1.04	1.20	1.59
Grand Total	7,349	62,470,220	8,500.50	1.17	1.42	1.54

*** The median of HSA specific nr sub-units SDs (1.94) is statistically different (Kolmogorov-Smirnov Test P <2.2e-16) than the median of HSA shared (3.37).

Fig. S6: Copy-number distribution of primate segmental duplications. A non-redundant set of human segmental duplications[20] were classified as lineage-specific or shared among the four primate species and the copy-number of each duplicon was estimated by the depth-of-coverage analysis (WSSD). The percentage of each category distributed across different copy-number bins is indicated. Lineage-specific duplications (colored histograms) show significantly fewer copies than shared duplications (in different intensities of grey). The copy number of every SD was calculated independently according to their species' depth of coverage.

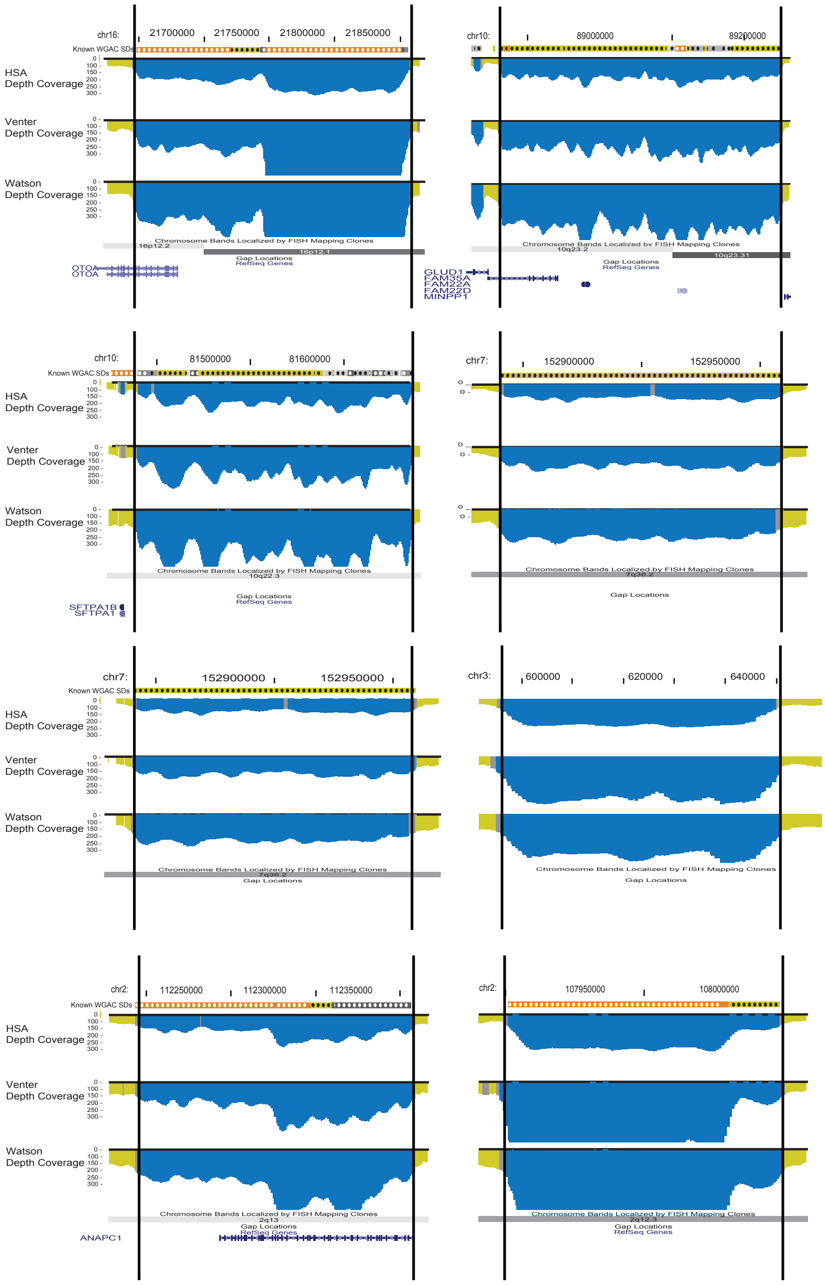


Fig. S7: WSSD duplication analysis of two human genomes. We performed the depth-of-coverage analysis of two human genomes (Venter/HuRef (Levy et al. 2007) and Watson (Wheeler et al. 2008)) and constructed two independent duplication maps for each to assess the extent of variation. We found that 95% of the duplication intervals (>20 kbp in length) were confirmed between these two genomes with the boundaries showing remarkable specificity. We depict 8 different intervals of the human assembly (build35) comparing the computationally predicted regions of duplication (blue) and unique sequence (yellow) with an assembly based analysis of human segmental duplications (WGAC analysis, top bar).