Figure S7. **Limited variation in the abundance of centromeric arrays is seen in populations of diverse ethnicity investigated from the 1000 genomes project**. Heat map representing the abundance of α-repeats in centromeric arrays, pericentromeric proviruses K111 and K222, and single copy genes (right Y-axis) determined *in silico* by BLAST analysis of Illumina sequences from diverse populations (X-axis). The $\log_2$ Z-score of each array is depicted by a gradient color scale on the left of the heat map. High and low copy numbers are thus indicated by blue and white, respectively. The analysis was performed by obtaining the number of sequence reads in each population that BLAST to ~100 bp of query sequence, allowing no more than 10 base pair mismatches and indels. Hierarchical clustering was performed to differentiate between clades of repeats based on abundance (left tree). The tree splits into two main branches, one indicating arrays in centromeric "cores," present in higher abundance, and "pericentromere" arrays that include K111 and K222 proviruses. The latter branch further splits into a branch representing single copy genes. No significant gene variation was seen in the single copy *TOP3A*, *CCR5*, *DEK*, or *ACTB* genes. Surprisingly, using our BLAST analysis, we did not retrieve significant matches for p82H, a repeat that is present in the centromeres of all human chromosomes, perhaps because p82H sequences have more mutations than the number allowed under our screening criteria. Analysis of pericentromere proviruses could only be performed on the 5' integration site area specific to K111/K222 in order to avoid sequence hits from other endogenous proviruses of the same HERV-K family that resemble the K111/K222 proviral genome by more than 95%. ACB: African Caribbean in Barbados, ASW: African Ancestry in Southwest US, CEU: European from Utah, CHS: Southern Han Chinese, China, CLM: Colombian in Medellin, Colombia, GBR: British in England and Scotland, GWD: Gambian in Western Division, The Gambia, GIH: Gujarati Indian in Houston,TX, IBS: Iberian populations in Spain, JPT: Japanese

in Tokyo, Japan, KHV: Kinh in Ho Chi Minh City, Vietnam, LWK: Luhya in Webuye, Kenya, MXL: Mexican Ancestry in Los Angeles, California, PEL: Peruvian in Lima, Peru, PUR: Puerto Rican in Puerto Rico, TSI: Toscani in Italy, and YRI: Yoruba in Ibadan, Nigeria.