SUPPLEMENTARY NOTE

Between-species differences in CNV genetic diversity

We observed a higher proportion of rare CNVRs in chimpanzees (209 out of 438 total CNVRs; 48%) than in humans (130 out of 353 total CNVRs; 37%). All but one of the chimpanzee individuals in our study are from a single "population" of chimpanzees, i.e., the western chimpanzee subspecies (*Pan troglodytes verus*). While Becquet et al. (2007) describe the presence of some population structure *within* western chimpanzees based on genome-wide microsatellite data, we would expect any population structure-related effects on CNV genetic diversity to be generally greater in our human sample, which is comprised of three distinct African populations. However, sampling more populations is expected to result in the identification of a higher proportion of rare alleles (Ptak and Przeworski 2002), the opposite of our observations for human and chimpanzee CNVRs.

Another possibility is that the higher proportion of rare CNVRs observed in chimpanzees simply reflects general demographic differences (e.g., in population growth) between our two species (and specifically, between western chimpanzees and sub-Saharan African humans). Indeed, based on intergenic region nucleotide sequence data, Fischer et al. (2006) report a slightly higher proportion of low frequency alleles in western chimpanzees compared to the human Hausa population (Cameroon). However, we cannot exclude factors that may also have contributed to our observation, such as different CNV mutation rates in humans and chimpanzees, different natural selection pressures on CNVs, or technical artifacts.

Overlap between CNVRs, CNDs, and loci from the Database of Genomic Variants (DGV)

Of the 353 human CNVRs we detected, 313 (89%) were found to overlap copy-number variable regions identified in previous human studies, as annotated in the Database of Genomic Variants (DGV, version hg18.v3, http://projects.tcag.ca/variation). However, based on a permutation analysis, 222 (63%) of these human CNVRs were expected to overlap CNVs from DGV at random. This may reflect the fact that the DGV now contains data from a diversity of previous CNV studies based on many different platforms with different effective resolutions. In consequence, even when the previously recorded CNVs (or those in the present study) are not false positives, overlapping CNVs may not reflect the detection of homologous events, especially if the inaccuracy of estimated boundaries leads to false conclusions of overlap.

Similarly, of the 92 fixed CNDs we detected, 27 (29.3%) did not overlap any DGV locus. This represents a considerable enrichment when compared to only 8 of 181 non-fixed CNDs (4.4%) which overlap human CNVRs from our set but no CNV from DGV (**Supplementary Table 1**), and this does not mean that all of the remaining 65 CNDs are not truly "fixed": the overlapping DGV loci may as well be false positives or not homologous. Further characterization of CND/CNV loci will be required to address this question.

CNV frequency distribution analyses

CNV mutation rates, and thereby neutral frequency distributions, may be dissimilar in SD compared to non-SD regions (as discussed in the manuscript, some SD regions of the genome may be subject to recurrent CNV genesis). Supporting this notion, we observed a higher proportion of rare variants for CNVRs that do not overlap SDs in either genome, in both humans (no SDs: 87 rare vs. 80 common CNVRs; overlap SDs: 43 rare vs. 143 common CNVRs;

Fisher's exact test; $P < 1x10^{-7}$) and chimpanzees (no SDs: 151 rare vs. 101 common; overlap SDs: 58 rare vs. 128 common; $P < 1x10^{-7}$). Therefore, we considered separately these CNV subsets, focusing on SD-overlapping CNVRs because in humans 177 of these variants intersect one or more genes, versus only 124 SD-absent CNVRs, providing larger sample sizes for individual Gene Ontology functional categories. However, there were only 9 SD-overlapping human CNVRs (8 in chimpanzees) mapped to intergenic regions (which reflects the strong enrichment of genes in SD regions of our genome (Cooper et al. 2007; Zhang et al. 2005)), obviating the use of any CNV subset as a neutral proxy for comparison to the gene-containing CNVR frequency distributions in a formal test of natural selection.

REFERENCES CITED

Becquet, C., N. Patterson, A.C. Stone, M. Przeworski, and D. Reich. 2007. Genetic structure of chimpanzee populations. *PLoS Genet* **3:** e66.

Cooper, G.M., D.A. Nickerson, and E.E. Eichler. 2007. Mutational and selective effects on copy-number variants in the human genome. *Nat Genet* **39:** S22-S29.

Fischer, A., J. Pollack, O. Thalmann, B. Nickel, and S. Paabo. 2006. Demographic history and genetic differentiation in apes. *Curr Biol* **16:** 1133-1138.

Ptak, S.E. and M. Przeworski. 2002. Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet* **18:** 559-563.

Zhang, L., H.H. Lu, W.Y. Chung, J. Yang, and W.H. Li. 2005. Patterns of segmental duplication in the human genome. *Mol Biol Evol* **22:** 135-141.