

Supplementary Figures

A Refined Analysis of the Neanderthal Sequences in Modern Humans with the Complete Reference Genome

Shen-Ao Liang^{1†}, Tianxin Ren^{2†}, Jiayu Zhang^{1†}, Jiahui He³, Xuankai Wang², Xinrui Jiang², Yuan He¹, Rajiv C. McCoy⁴, Qiaomei Fu^{5,6}, Joshua M. Akey⁷, Yafei Mao^{2,8*}, Lu Chen^{1*}

* Correspondence: yafmao@sjtu.edu.cn (Y.M.); lu_chen@fudan.edu.cn (L.C.)

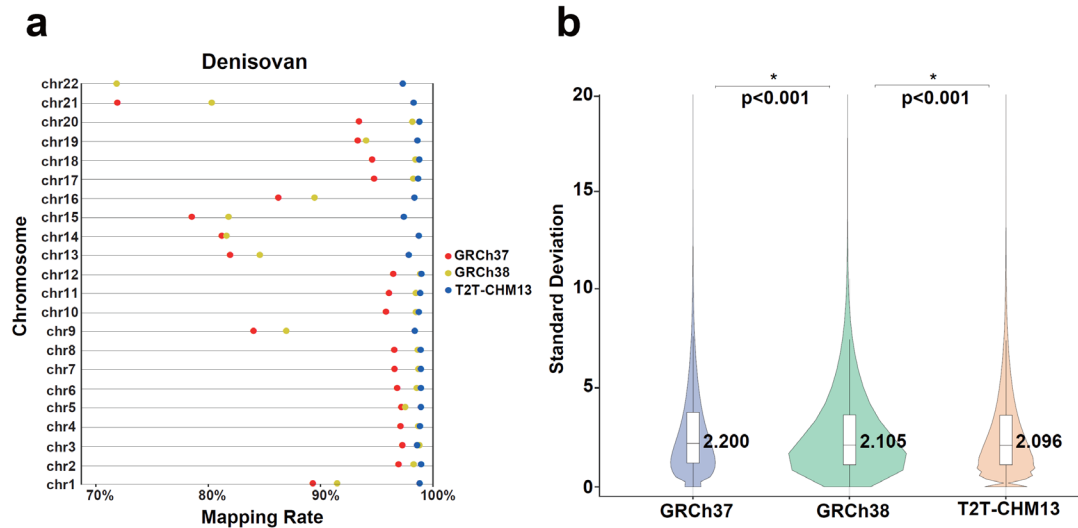


Fig. S1. Comparison of Denisovan statistics across three reference genomes. (a) Denisovan read mapping rate in unmasked genomic regions among GRCh37, GRCh38 and T2T-CHM13. (b) Violin plots showing the standard deviation (s.d.) of Denisovan read depth in unmasked genomic regions among GRCh37, GRCh38 and T2T-CHM13. A significant reduction in the s.d. of read coverage is observed for GRCh38 compared to GRCh37 (p-value $< 1e-323$) and for T2T-CHM13 compared to GRCh38 (p-value = $2.0057e-160$).

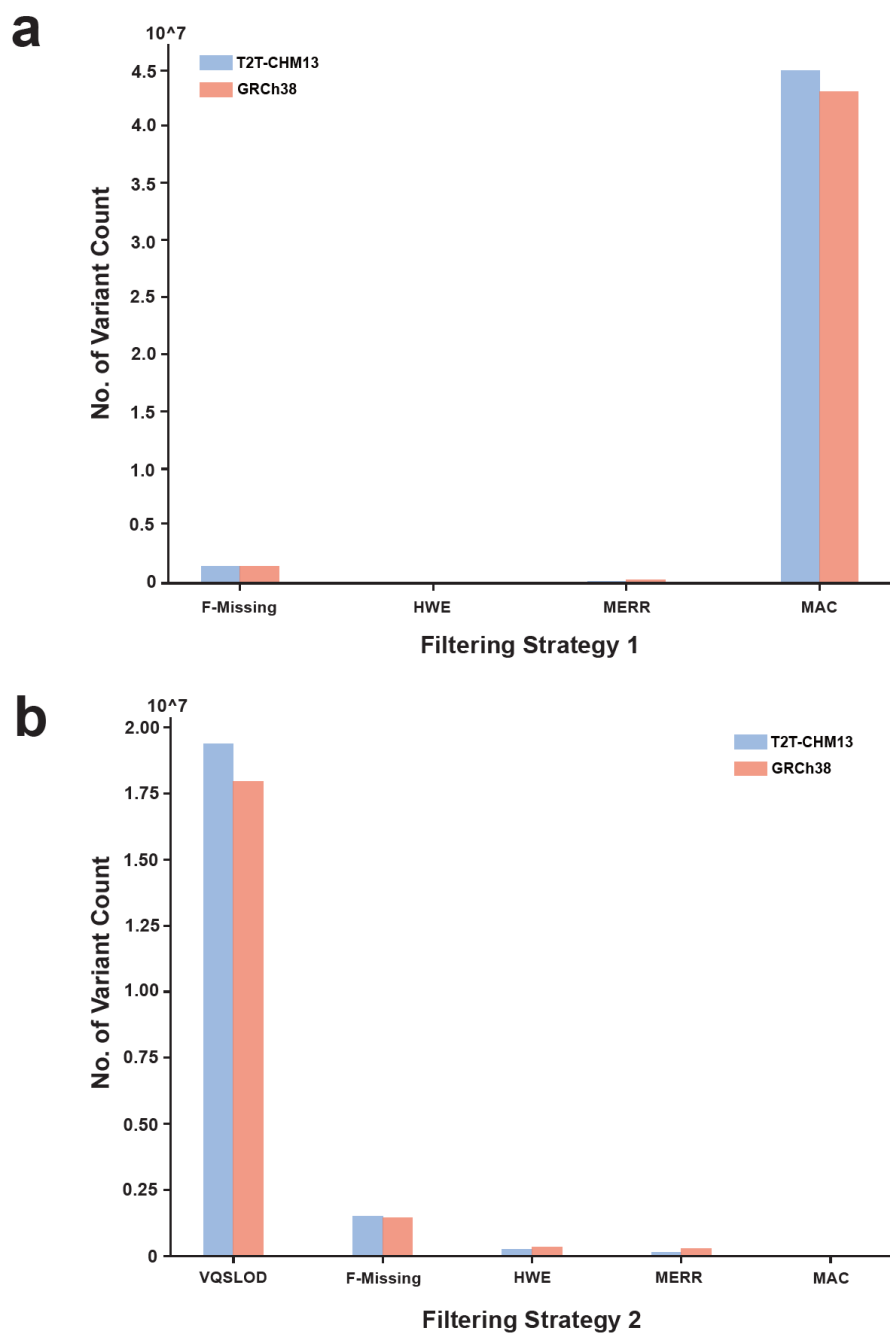


Fig. S2. Count of biallelic variants excluded by each filtering criterion in pre-phasing filtering strategies. Bar plots showing count of biallelic variants excluded by each filtering criterion in two pre-phasing filtering strategies in T2T-CHM13 and GRCh38. (a) Strategy 1. (b) Strategy 2.

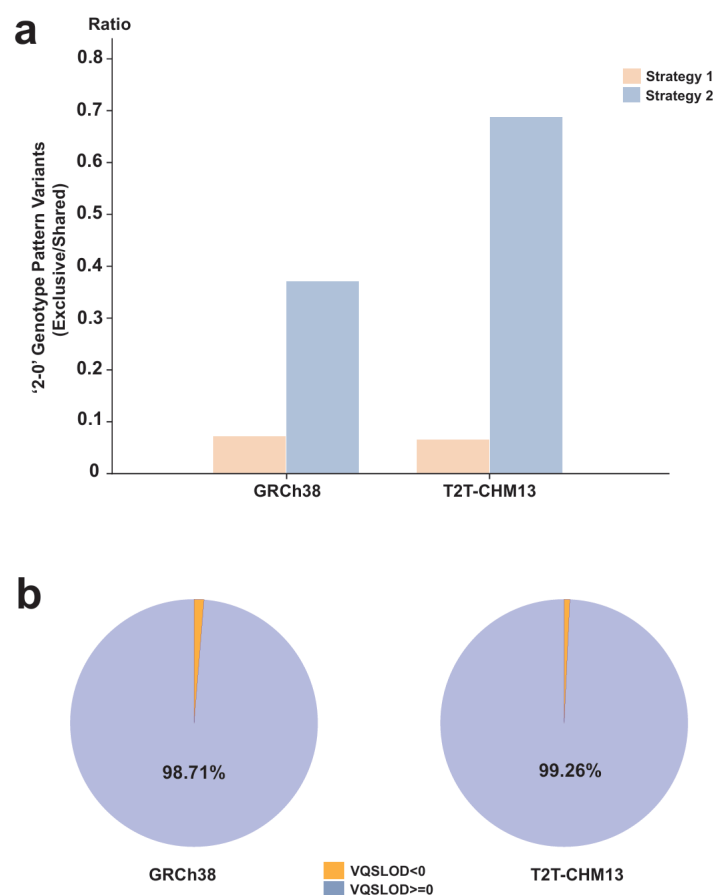


Fig. S3. ‘2-0’ genotype pattern variants under two pre-phasing strategies. (a) The ratio of exclusive ‘2-0’ genotype pattern variant count under each strategy and two-strategies-shared ‘2-0’ genotype pattern variant count in GRCh38 and T2T-CHM13. (b) Pie charts showing the ratio of the ‘2-0’ genotype pattern variants caused by VQSLOD criterion of Strategy 2 in GRCh38 (left) and T2T-CHM13 (right).

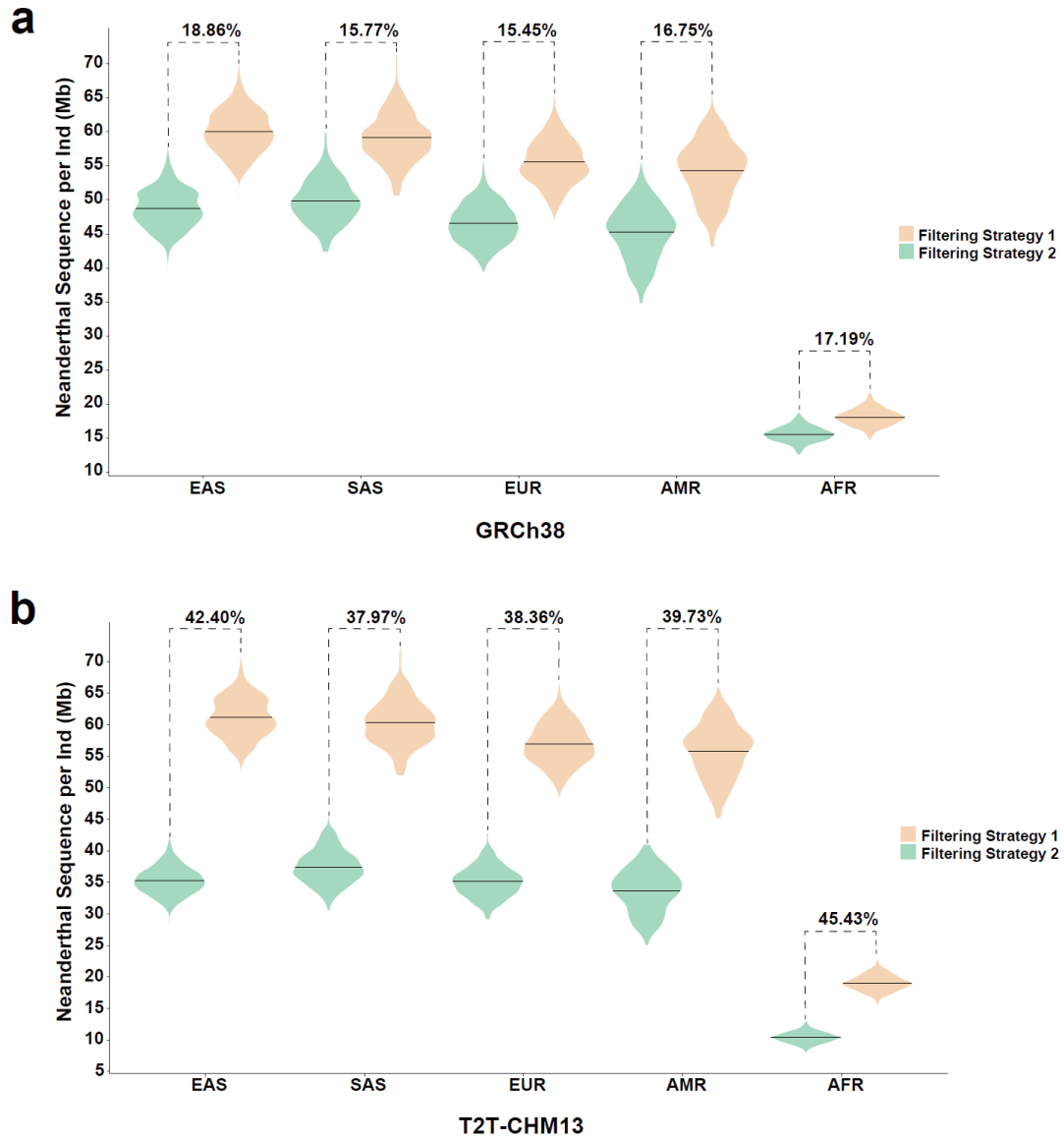


Fig. S4. Comparison of Neanderthal introgression detected under different pre-phasing filtering strategies. Violin plots showing Neanderthal introgression bias per individual under two different pre-phasing filtering strategies in (a) GRCh38 and (b) T2T-CHM13.

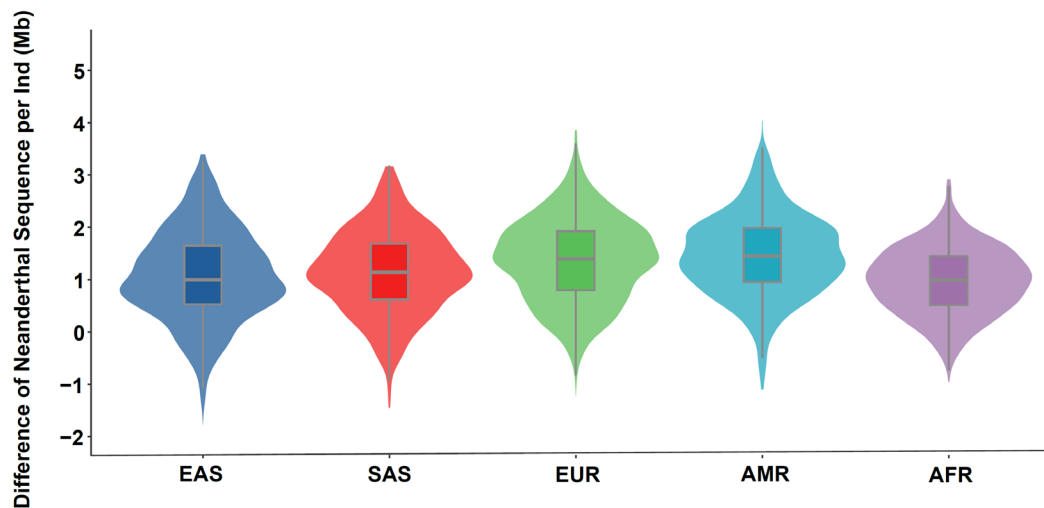


Fig. S5. Difference of Neanderthal introgression per individual between T2T-CHM13 and GRCh38. Violin plots showing difference (Mb) of Neanderthal introgressed sequences per individual identified between T2T-CHM13 and GRCh38 in five populations from the 1000 Genomes Project (1KGP).