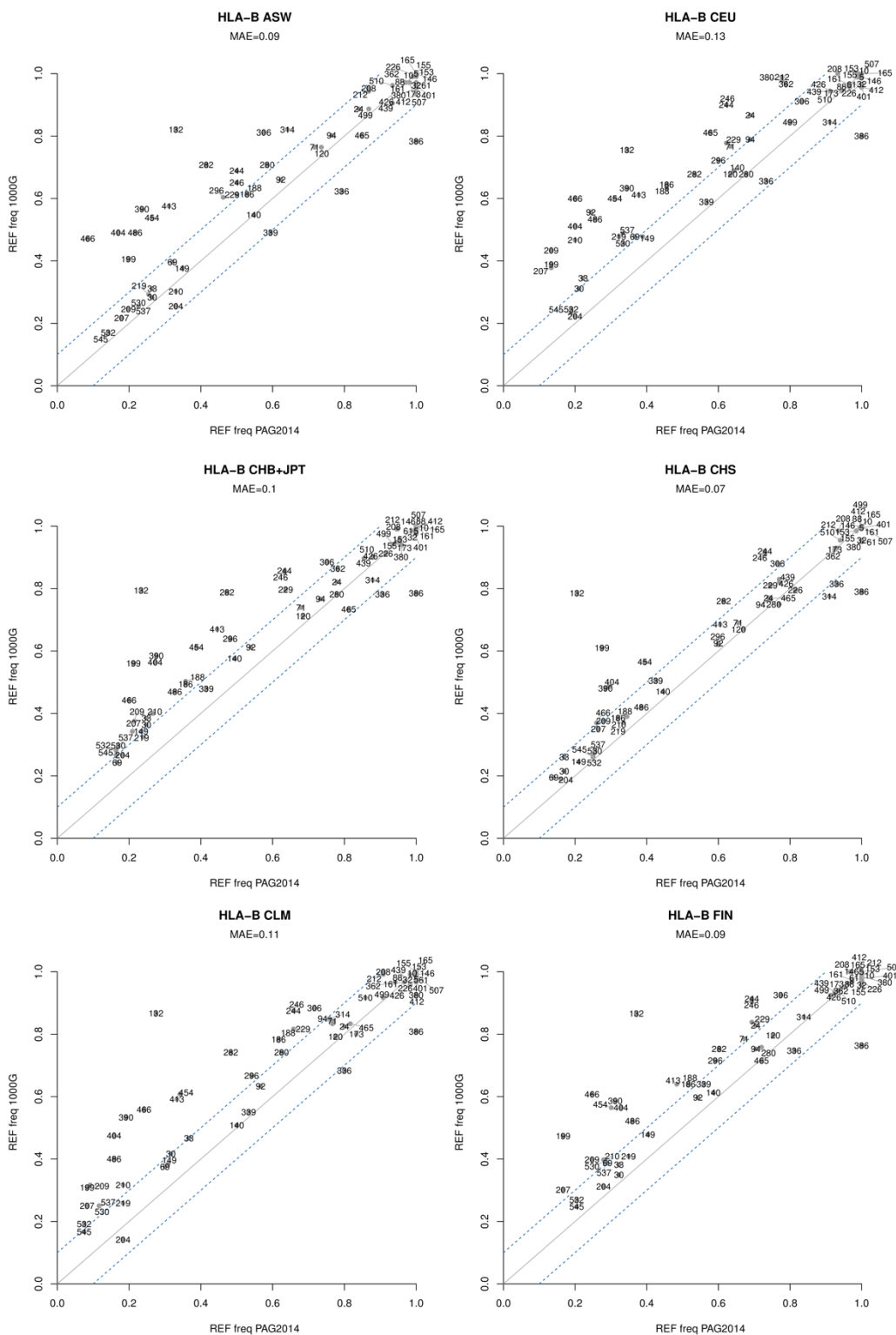
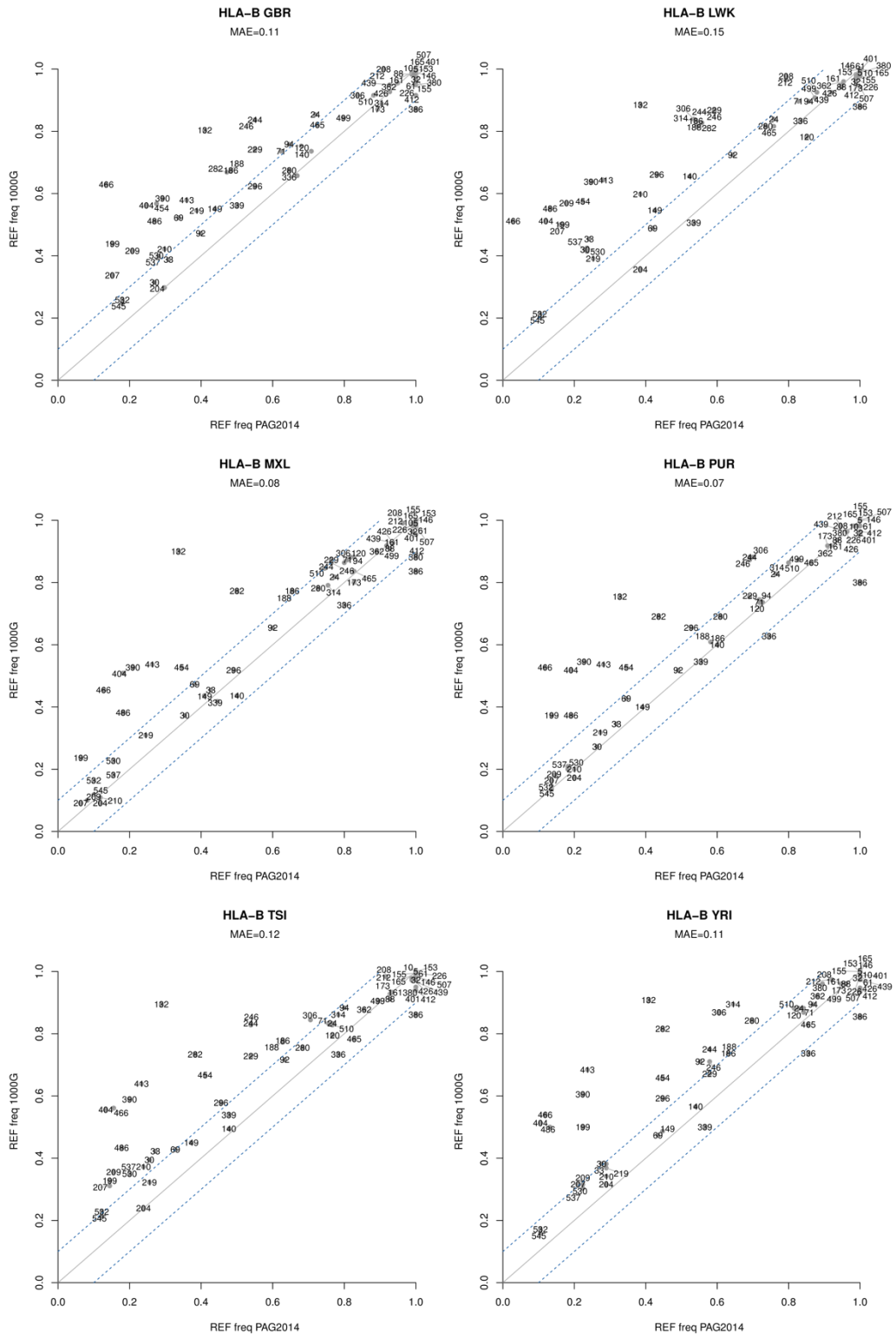


Figure S4





**Figure S4** Reference allele frequency per population and per site in the *HLA-B* gene in the 1000 Genomes (1000G; y-axis) and Sanger sequencing (PAG2014; x-axis) datasets. Dashed lines indicate a  $\pm 0.1$  deviation from the expected frequency (as estimated from PAG2014 dataset). MAE (mean absolute error) defined in Methods. Numbers indicate site position in ARS exons sequence.