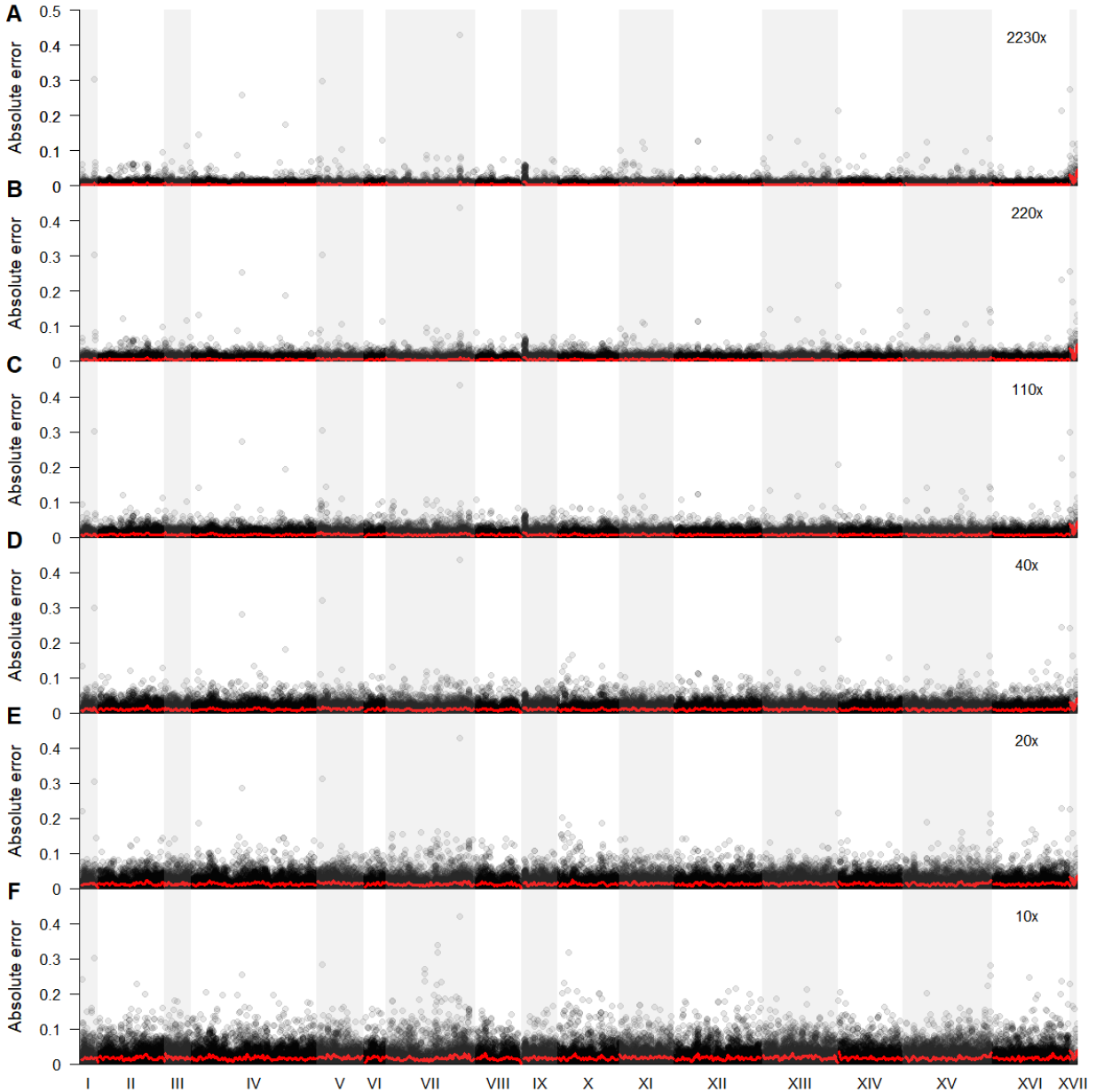

Figure S4. Shown is the absolute error rate between haplotype and SNP frequencies at sites with private SNPs for 18F12v2. Panels A-F represent 100%, 10%, 5%, 2%, 1%, and 0.5% of the full dataset, which approximately corresponds to the coverages shown in the upper right of each panel. The red line represents a kernel regression run using the ksmooth() function in R with kernel set at 'normal' and bandwidth set at 20000, approximating the local expected error rate. Despite clear outliers (that occur irrespective of coverage), haplotype frequency estimates are accurate on average.