



**S. Figure 1.** For a sample of 1,000 SNPs (in the 86k set), there is a significant linear relationship between SNP minor allele frequency and percentage of missing genotypes ( $P < 2 \times 10^{-16}$ ).