S2 Type-1 Diabetes Dataset

Due to the pooling-based sequencing method employed, the data on Type-1 diabetes available to us consists of frequency counts at 178 rare variants (frequency less that ~ 0.03). Since all methods discussed are based on having complete genotype data for cases and controls, it is necessary to reconstruct the actual genetic data for the 480 cases, and 480 controls. We have generated 100 such datasets; for each such dataset, the observed frequency count at each variant is fixed at the value observed in the original study. We have assumed that rare variants are independent (this is a reasonable assumption, given that the variants are rare).