

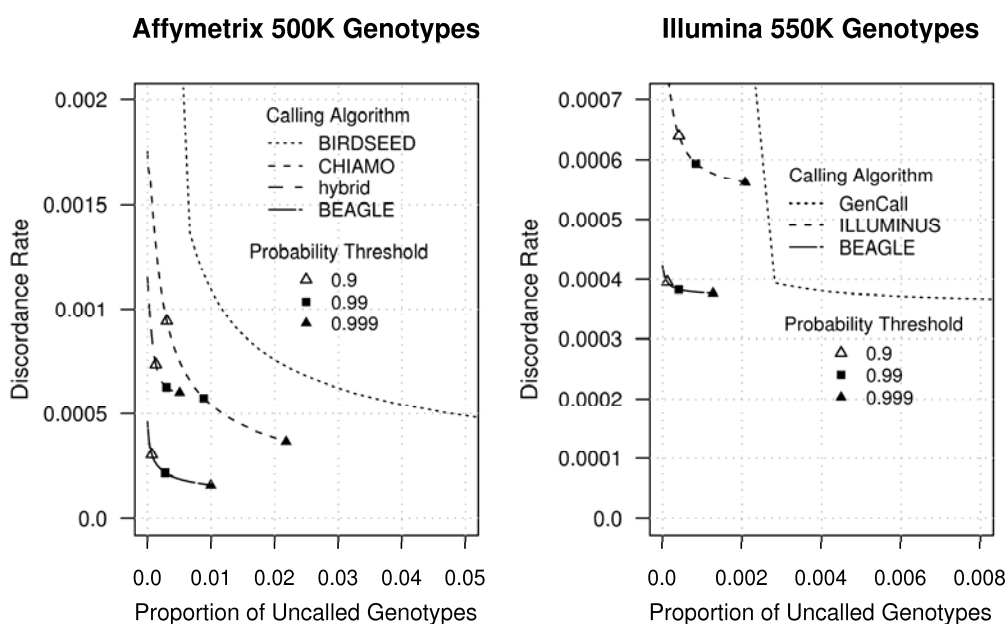
Supplemental Data

AJHG, Volume 85

Simultaneous Genotype Calling and Haplotype Phasing Improves Genotype Accuracy and Reduces False-Positive Associations for Genome-Wide Association Studies

Brian L. Browning and Zhaoxia Yu

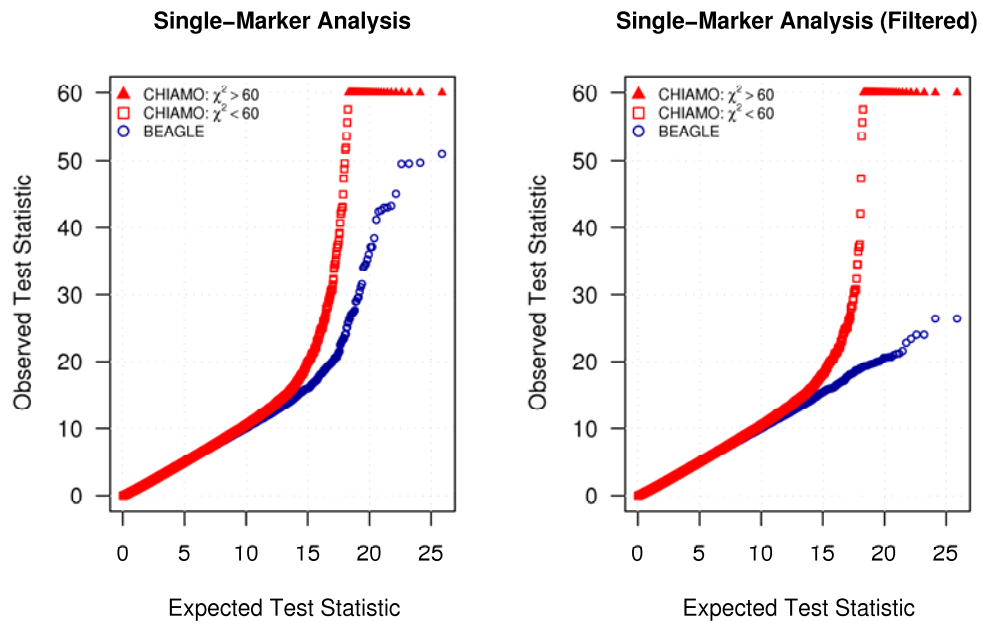
Figure S1. Genotype Discordance and Missing Data Rates Using GenCall and CHIAMO Reference Genotypes



Discordance rates for genotype calls for autosomal Affymetrix 500K chip data (left panel) and autosomal Illumina 550K chip data (right panel) are computed using high confidence CHIAMO or GenCall genotype calls from the alternate platform. Discordance rates for Illumina genotypes are computed using CHIAMO genotypes with ≥ 0.99995 posterior probability. Discordance rates for Affymetrix genotypes are computed using GenCall genotypes with ≥ 0.39995 GenCall score. The proportion of missing genotypes and the discordance rate for called genotypes are computed for each possible calling threshold for

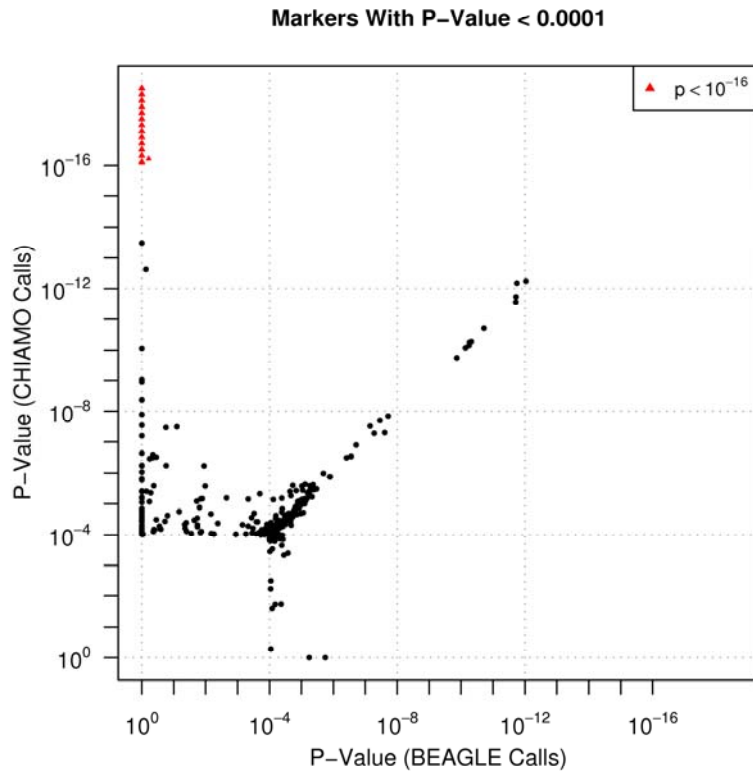
each method. The discordance and missing data rates corresponding to calling thresholds of 0.9, 0.99, and 0.999 posterior genotype probability are shown for the genotype calling methods that report genotype probabilities.

Figure S2. Quantile-Quantile Plots for Single-Marker Analysis of Type 2 Diabetes



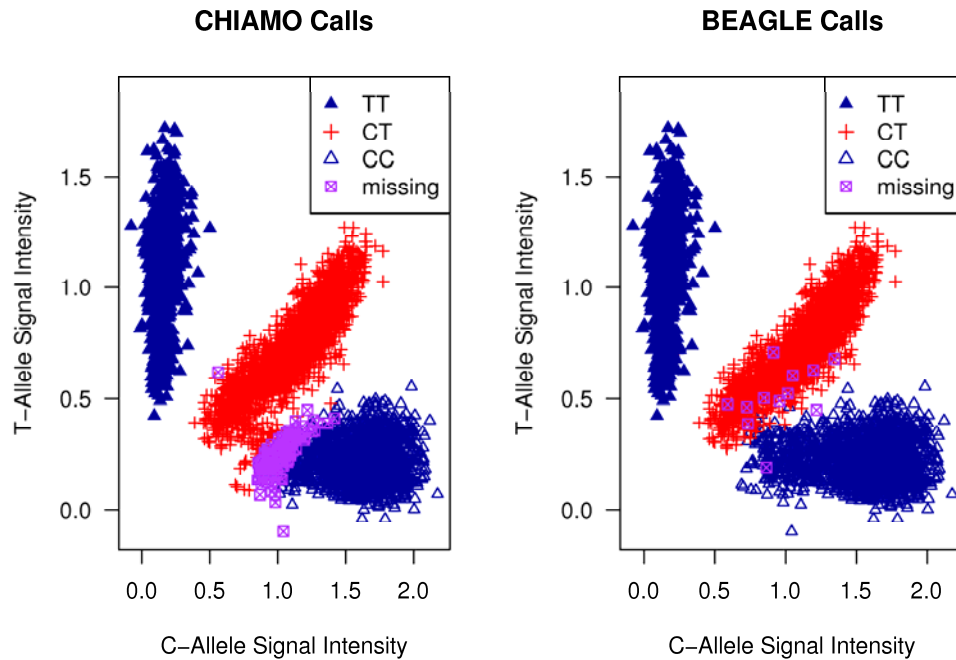
Expected and observed association chi-square test statistics from analysis of CHIAMO genotype calls and BEAGLE genotype calls of WTCCC type 2 diabetes and control data. An allelic test statistic and three genotypic test statistics, corresponding to dominant, overdominant, and recessive models, are computed for each marker. The right panel excludes 66 markers in three regions that WTCCC described as having replicated association to type 2 diabetes. The three regions with excluded markers are chromosome 6p22 (20.63-20.84 Mb), chromosome 10q25 (114.71-114.81 Mb), and chromosome 16q12 (52.36-52.41 Mb). Positions are in NCBI Build 35 coordinates.

Figure S3. P-Values from Single-Marker Analysis of WTCCC Type 2 Diabetes and Control Data



The minimum P-value from an allelic trend test, and 3 genotypic tests (for dominant, overdominant, and recessive models) was calculated for each marker for CHIAMO and BEAGLE genotype calls. The P-values from CHIAMO calls and BEAGLE calls are plotted using a log scale for all markers with minimum P-value < 0.0001 for one or both genotype calling methods. P-values for markers that were excluded by data quality filters for CHIAMO calls but not by data quality filter for BEAGLE calls are plotted along the line $y = 1$. P-values for markers that were excluded by data quality filters for BEAGLE calls but not by data quality filters for CHIAMO calls are plotted along the line $x = 1$.

Figure S4. Allele Signal Intensities and Genotype Calls for Marker rs5015480



Allele signal intensities, CHIAMO genotype calls (left panel), and BEAGLE genotype calls (right panel) for a marker associated with type 2 diabetes (rs5015480) for 4862 individuals from the T2D, 58BC, and NBS cohorts that were genotyped on the Affymetrix 500K chip and that passed genome-wide quality control filters (see Material and Methods). In the left panel, 123 genotypes with CHIAMO posterior probability <0.90 are labelled as missing. In the right panel, 33 genotypes with BEAGLE posterior probability <0.97 are labelled as missing. The CHIAMO and BEAGLE genotype calls have 98.2% and 99.9% concordance respectively with genotype calls for the 58BC cohort from Illumina 550K chip data.