The design of GWAS has improved in recent years with larger samples and multi-tiered replication. Minimal errors in such large-scale studies can introduce bias that inflates both Type I and Type II error rates, making adequate quality control essential to ensure reliability of GWAS results.

Recent research has shown that discordant GWAS results can arise due to differences in both batch size (the number of samples called simultaneously) and batch composition (combining or separating cases and controls and/or different populations in batch sets).

We evaluated the effects of different batch sizes and case-control compositions on the outcome of GWAS results using previously published data from the Wellcome Trust Case Control Consortium (WTCCC) using the following two different genotyping algorithms: Bayesian Robust Linear Model with Mahalanobis distance classifier (BRLMM) and the Bayesian Hierarchical Mixture Model (CHIAMO).

**Methods**

**Data source:**
Data were obtained from the WTCCC. The Affymetrix GeneChip Human Mapping 500K array genotypes data was downloaded from 1991 coronary artery disease (CAD) cases and 1000 UK Blood Service Controls.

**Genotyping calling algorithms:**
BRLMM and CHIAMO algorithms were used to call for the Nip and Sty arrays separately.

**BRLMM (Affymetrix Power Tools v1.0.2):**
BRLMM is a multiplex algorithm to assign calls using an optimized Mahalanobis distance (d1) to a genotype cluster center (AA, AB, BB). A confidence score of the call (d1-d2, where d2 is the distance to the nearest closest cluster) determines if a call is set as missing. A Bayesian step employs a single-chip algorithm on a SNP sample to estimate priors for cluster centers and variances, improving call rates and accuracy.

**CHIAMO (v0.2.1):**
The CHIAMO algorithm, developed for the original WTCCC analysis, assigns calls using a Bayesian Hierarchical 4-class (AA, AB, BB, null mixture model). The optimal hyper-parameters are found by maximizing the posterior model distribution, given 12 random starting points. The algorithm supports simultaneous calling from all individuals and multiple cohorts.

**Batch Effect Configurations:**
Samples were separated into the following 5 batches:
- S500: 500 cases x 4 and 500 controls x 3
- S2000: 1991 cases x 1 and 1500 controls x 1
- C500: 285 cases and 215 controls combined x 7
- C2000: 995 cases and 750 controls combined x 2
- C3500: all 3491 samples combined x 1

1991 cases, therefore batches may have contained more or less samples for that scenario.

**Quality control and association testing:**
Analyses were conducted using JMP Genomics Statistical Discovery Software from SAS Institute or PHStat (v1.06).

Samples were excluded if the SNP call rate was less than 95% and if the minor allele frequency (MAF) was less than 1%, or the call rate was less than 95%, 2) p<5x10^-7 for a d.f. trend test for proportion of missing data between cases and controls, and 3) if 4-class (AA, AB, BB, null mixture model) for control Hardy-Weinberg Equilibrium (HWE).

A stringent genotyping call was also performed for SNPs that were called in all previous QC steps as well as had a call rate no less than 99%.

SNP association with CAD status was performed using Cochran-Armitage trend test [19] for additive allele effects for each of the five datasets: p-values for a 1 d.f. test less than 5x10^-7 were considered significant.

**Discussion**

**Previous studies of GWAS data processing have demonstrated the following:**

- Genotype clusters differ positionally for cases and controls, resulting in ambiguous genotype calls when samples are combined and overdispersion of association test statistics. Stringent QC and confidence limits can lead to bias due to non-independence in genotypes set to missing. (Clayton et al)
- Allowing different priors for frequencies in genotype clusters calling algorithms between cases and controls conflicts with the null hypothesis of no association with genetic markers and case-control status. (Plagnol et al)
- Differential bias due to genotype errors can be remedied with stringent QC (namely call rate), but this results in significant data loss. ( Miyagawa et al)
- Processing different batch sizes may affect sample and SNP call rates. Ancestral batch composition factored in the results could be propagated to simulated association testing results. A primary recommendation was to use ancestrally homogeneous, larger batch sizes. (Hong et al)

Using Affymetrix 500K array data obtained from 3,491 individuals in the WTCCC CAD GWAS, we determined that batch size and case/control composition affect common QC steps and association testing, which produces differences in the lists of SNPs found to be significant in GWAS. Our studies show that for BRLMM and CHIAMO, using larger batch sizes of combined case and control samples for genotype calling yields more conservative, yet more consistent association testing results.

**Recommendations:**
- Use large batch sizes for genotype calling
- Pool cases and controls when case/control sample collection and processing procedures are uniform
- Conduct sensitivity analysis with stringent QC thresholds