RESEARCH ARTICLE

Evaluating variations of genotype calling: a potential source of spurious associations in genome-wide association studies

HUIXIAO HONG[1]*, ZHENQIANG SU[1], WEIGONG GE[1], LEMING SHI[1], ROGER PERKINS[1], HONG FANG[2], DONNA MENDRICK[1] and WEIDA TONG[1]

[1]*Division of Systems Toxicology, National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA*

[2]*Z-Tech Corp, ICF International Company at National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA*

Abstract

Genome-wide association studies (GWAS) examine the entire human genome with the goal of identifying genetic variants (usually single nucleotide polymorphisms (SNPs)) that are associated with phenotypic traits such as disease status and drug response. The discordance of significantly associated SNPs for the same disease identified from different GWAS indicates that false associations exist in such results. In addition to the possible sources of spurious associations that have been investigated and discussed intensively, such as sample size and population stratification, an accurate and reproducible genotype calling algorithm is required for concordant GWAS results from different studies. However, variations of genotype calling of an algorithm and their effects on significantly associated SNPs identified in downstream association analyses have not been systematically investigated. In this paper, the variations of genotype calling using the Bayesian Robust Linear Model with Mahalanobis distance classifier (BRLMM) algorithm and the resulting influence on the lists of significantly associated SNPs were evaluated using the raw data of 270 HapMap samples analysed with the Affymetrix Human Mapping 500K Array Set (Affy500K) by changing algorithmic parameters. Modified were the Dynamic Model (DM) call confidence threshold (threshold) and the number of randomly selected SNPs (size). Comparative analysis of the calling results and the corresponding lists of significantly associated SNPs identified through association analysis revealed that algorithmic parameters used in BRLMM affected the genotype calls and the significantly associated SNPs. Both the threshold and the size affected the called genotypes and the lists of significantly associated SNPs in association analysis. The effect of the threshold was much larger than the effect of the size. Moreover, the heterozygous calls had lower consistency compared to the homozygous calls.

[Hong H., Su Z., Ge W., Shi L., Perkins R., Fang H., Mendrick D. and Tang W. 2010 Evaluating variations of genotype calling: a potential source of spurious associations in genome-wide association studies *J. Genet.* **89**, 55–64]

Introduction

The International HapMap project determined genotypes over 3.1 million common SNPs in human populations and computationally assembled them into a genome-wide map of SNP-tagged haplotypes (The International HapMap Consortium 2005, 2007). Concurrently, advances in high-throughput SNP genotyping technology enabled the simultaneous genotyping of hundreds of thousands of SNPs. These advances combined to make genome-wide association studies (GWAS) a feasible and a promising research field for associating genotypes with various disease susceptibilities and health outcomes.

Recently, GWAS was successfully applied to identify common genetic variants associated with a variety of phenotypes (Klein *et al*. 2005; Duerr *et al*. 2006; Smyth *et al*. 2006; Buch *et al*. 2007; Cargill *et al*. 2007; Easton *et al*. 2007; Frayling *et al*. 2007; Grupe *et al*. 2007; Gudmundsson *et al*. 2007; Hampe *et al*. 2007; Hunter *et al*. 2007; Raelson *et al*. 2007; Rioux *et al*. 2007; Saxena *et al*. 2007; Scott *et al*. 2007; Sladek *et al*. 2007; Steinthorsdottir *et al*. 2007; Todd *et al*. 2007; Tomlinson *et al*. 2007; van Heel *et al*. 2007; Wellcome Trust Case Control Consortium 2007; Winkelmann *et*

*For correspondence. E-mail: huixiao.hong@fda.hhs.gov.
The views presented in this article do not necessarily reflect those of the US Food and Drug Administration.

Keywords. genotype calling; genome-wide association studies; missing call rate; calling algorithm; spurious association.

al. 2007; Yeager *et al*. 2007; Zanke *et al*. 2007; Zeggini *et al*. 2007; Arking *et al*. 2008; Butcher *et al*. 2008; Gold *et al*. 2008; Kayser *et al*. 2008; Uda *et al*. 2008; Yang *et al*. 2008). These findings are valuable for scientists to elucidate the allelic architecture of complex traits in general. However, replication studies of GWAS showed that only a small portion of significantly associated SNPs in the initial GWAS results can be reproduced in people within the same population. For example, GWAS studies in patients with type 2 diabetes mellitus have found poor replication rates in the range of 0–13% (Scott *et al*. 2007; Sladek *et al*. 2007; Steinthorsdottir *et al*. 2007; Zanke *et al*. 2007; Moore *et al*. 2008), demonstrating the many false positives in current GWAS. Moreover, the lists of significantly associated SNPs identified in different GWAS for the same disease, such as type 2 diabetes mellitus, were quite different across studies. It is obvious that there are potentially type I (false positive) and type II (false negative) errors in GWAS results, limiting the potential for early application to personalized medicine and nutrition.

A genotype calling algorithm is a set of mathematical transformations that are used to convert the raw intensity data to the genotypes for the downstream association analysis. Highly accurate and reproducible genotype calls are paramount for the success of GWAS since errors introduced in genotype calls can lead to inflation of type I and type II errors. Genotyping error, especially if occurring differentially between cases and controls, are an important cause of spurious associations and should be carefully examined and corrected (Moskvina *et al*. 2006). A number of quality control features are advocated to be used on both a per-sample and a per-SNP basis. One of the fundamental questions in GWAS is how consistent the genotype calls are when obtained upon altering the parameters within an algorithm.

The Affy500K platform (A flymetrix, Sata Clana, USA) was used in many GWAS (Frayling *et al*. 2007; Saxena *et al*. 2007; Wellcome Trust Case Control Consortium 2007). The genomic DNA for one of the arrays is cleaved with the *Nsp*I restriction enzyme and ~262,000 SNPs are interrogated. The second array uses *Sty*I cleaved genomic DNA and ~238,000 SNPs are analysed. Raw data (CEL files) obtained from Affy500K are usually evaluated with the calling algorithm Bayesian Robust Linear Model with Mahalanobis distance classifier (BRLMM) (Affymetrix 2006) embedded in the Affymetrix software package. There are several parameters that need to be set for conducting genotype calling using BRLMM. The algorithm first derives an initial guess for each SNP's genotype using the DM algorithm (Di *et al*. 2005) and then analyses across SNPs to identify cases of nonmonomorphism. This subset of non-monomorphic SNPs is then used to estimate a prior distribution on cluster centres and variance–covariance matrices. This subset of SNP genotypes is revisited, and the clusters and variances of the initial genotype guesses are combined with the prior information of the SNPs in an *ad hoc* Bayesian procedure to derive a posterior estimate of cluster centres and variances. Genotypes of SNPs are called according to their Mahalanobis distances from the three cluster centres and confidence scores are assigned to the calls. The parameters that specify the confidence threshold for the DM algorithm with which to seed clusters (default value, 0.17) and the number of probe sets randomly selected for determining prior (default value, 10,000), influence on the prior distribution on cluster centers and variance-covariance matrices. Therefore, it is important to know whether changing values of the parameters causes inconsistent genotype calls and discordant lists of significantly associated SNPs in GWAS. However, to our knowledge, there are no systematic studies to examine variations in BRLMM parameters and how these may influence the generation of spurious associations in GWAS.

To assess whether the variation of genotype calling of BRLMM is a potential source of type I and type II errors in GWAS, we analysed how modifications of the threshold and size algorithmic parameters in BRLMM affect its ability to consistently call the 270 samples from the International HapMap project. To further examine whether the small discordance in genotypes is a possible source of spurious associations in GWAS results, we assessed whether this difference propagated to the list of significantly associated SNPs identified in the downstream analysis.

Materials and methods

Raw data

The raw data (CEL files) from the 270 HapMap samples profiled on the Affy500K were downloaded from the International HapMap project website (http://www.hapmap.org/downloads/raw_data/affy500k/). The CEL file format was described on Affymetrix's developer pages (http://www.affymetrix.com/Auth/support/developer/fusion/file_formats.zip). The file name indicated the population code (CEU/YRI/CHB+JPT), the sample identifier (e.g., NA12345), followed by the Affymetrix array type (based on restriction enzyme name: Nsp or Sty). Three population groups composed the data set and each group contained 90 samples: CEU had 90 samples from Utah residents with ancestry from northern and western Europe; CHB+JPT had 45 samples from Han Chinese in Beijing, P. R. China, and 45 samples from Japanese in Tokyo, Japan; YRI had 90 samples from Yoruba in Ibadan, Nigeria.

Quality of the raw data

The quality of the raw data from the Affy500K was assessed using DM (Di *et al*. 2005) before genotype calling by BRLMM. DM is a single array based algorithm; it processes one CEL file at a time in a multiple CEL file batch and statistically assesses experimental qualities with a numerical score between 0 and 100. A high QC (quality control) number means high quality of the experiment (CEL file).

Genotype calling by BRLMM

All experiments of genotype calling by BRLMM reported in this paper were conducted using apt-probeset-genotype of Affymetrix Power tools 1.8.5. Affymetrix Power tools (APT) contains a set of cross-platform command line programs that implement algorithms for analysing and working with Affymetrix GeneChip®arrays. These programs are available on the Affymetrix website (http://www.affymetrix.com/support/developer/powertools/index.affx). APT programs are intended for 'power users' who prefer programs that can be utilized in scripting environments and are sophisticated enough to handle the complexity of extra features and functionality. The function of apt-probeset-genotype in APT is an application for making genotype calls using SNP Arrays (100K, 500K, genome-wide SNP Arrays 5.0 and 6.0). BRLMM is one of the genotype calling algorithms implemented in this function, and enables many parameters to be changed by a user. For the study reported here, all the parameters, except as noted in the narrative, were set to the default values recommended by Affymetrix. The chip description files (cdf) for both Nsp and Sty chips of Affy500K, as well as files for defining SNPs on chromosome X, were also used before genotype calling. They were downloaded from Affymetrix website. Nsp and Sty CEL files were genotype-called separately.

In previous work, we assessed calling batch effect and found that uniform and large batch sizes with homogenous samples should be used to make genotype calls for GWAS (Hong *et al*. 2008). Therefore, three batches were used to make genotype calls; each used 90 samples from one of the three population groups.

Comparing genotype calling results

In each of the experiments reported here, the genotype calling results by BRLMM from using different thresholds and sizes were first merged using a set of in-house programs written in C^{++}. When merging the calling results, genotypes of SNPs in Nsp and Sty chips of the same samples were merged followed by assembling together all genotypes of all of the 270 HapMap samples. Thereafter, overall call rates for each of the experiments, missing call rates of individual samples and SNPs in each of the experiments, and concordant calls between experiments were calculated and exported as tab-delimited text files using the in-house programs written in C^{++}. Comparison of calling results was done using the *R* package.

Paired two samples *t*-test in *R* package (*t*-test) was used to statistically test the alternative hypothesis that missing call rates on samples or SNPs between two calling experiments are different.

Association analysis

In order to study the propagation of effects induced by algorithmic parameters to the significantly associated SNPs, all genotype calling results of the raw data of 270 HapMap samples using BRLMM with different thresholds and sizes were analysed using chi-square statistics test for associations between the SNPs and the case–control mimics.

Prior to association analysis, quality control (QC) of the calling results was conducted to remove markers and samples with low quality. For each of the calling results, call rate of 90% was used to remove SNPs. Minor allele frequency was used to filter SNPs and its cut-off was set to 0.01. Departure from Hardy–Weinberg equilibrium (HWE) was checked for all SNPs. The *P* value of chi-square test for Hardy–Weinberg equilibrium was calculated for all SNPs first and then the *P* values were adjusted for multiple tests using the Benjamini and Hochberg false discovery rate (FDR) (Benjamini and Hochberg 1995). FDR of 0.01 was set as the cut-off for HWE test.

To mimic 'case–control' in GWAS for the genotype calling results, each of the three population groups (European, African and Asian) was assigned as 'case' while the other two were used as a 'control'. This formed a data set for association analysis for identifying the SNPs significantly associated with the 'case' population group.

In the association analysis, a 2×3 contingency table (genotypic association) and a 2×2 contingency table (allelic association) were generated for each SNP and a case–control mimic. Then a chi-square statistics test was applied on the contingency tables to calculate the *P* values for measuring the statistical significance of the association between the testing SNP and the corresponding case–control mimic. After raw *P* values for all SNPs in a data set were calculated, Bonferroni correction was applied for *P* value adjustment. Lastly, a criterion of Bonferroni-corrected *P* value less than 0.01 was used to identify the significantly associated SNPs.

Results

Effect of changing the threshold parameter

Confidence thresholds of 0.17, 0.3, 0.45, and 0.6 were used to derive the initial genotypes of SNPs for estimating prior distribution in BRLMM. The overall missing call rates, defined as the proportion of missing to the total number of calls (successful plus missing calls), using thresholds of 0.17, 0.30, 0.45, and 0.60 were 0.5154%, 0.5312%, 0.6423%, and 0.7704%, respectively. Thus, as more confident DM calls were used to estimate the prior distribution for BRLMM, the missing calls were reduced. However, overall missing call rates are not sufficiently informative to assess their distribution in the data set. Therefore, the effect was compared using one-against-one comparisons of the distributions of missing call rates on individual samples and SNPs.

The comparisons of missing call rates of individual SNPs and samples using thresholds of 0.17, 0.30, 0.45, and 0.60 are given in the scatter plots of figures 1,A&B, respectively. The Pearson correlation coefficients of corresponding comparisons were calculated and are shown on the top of the

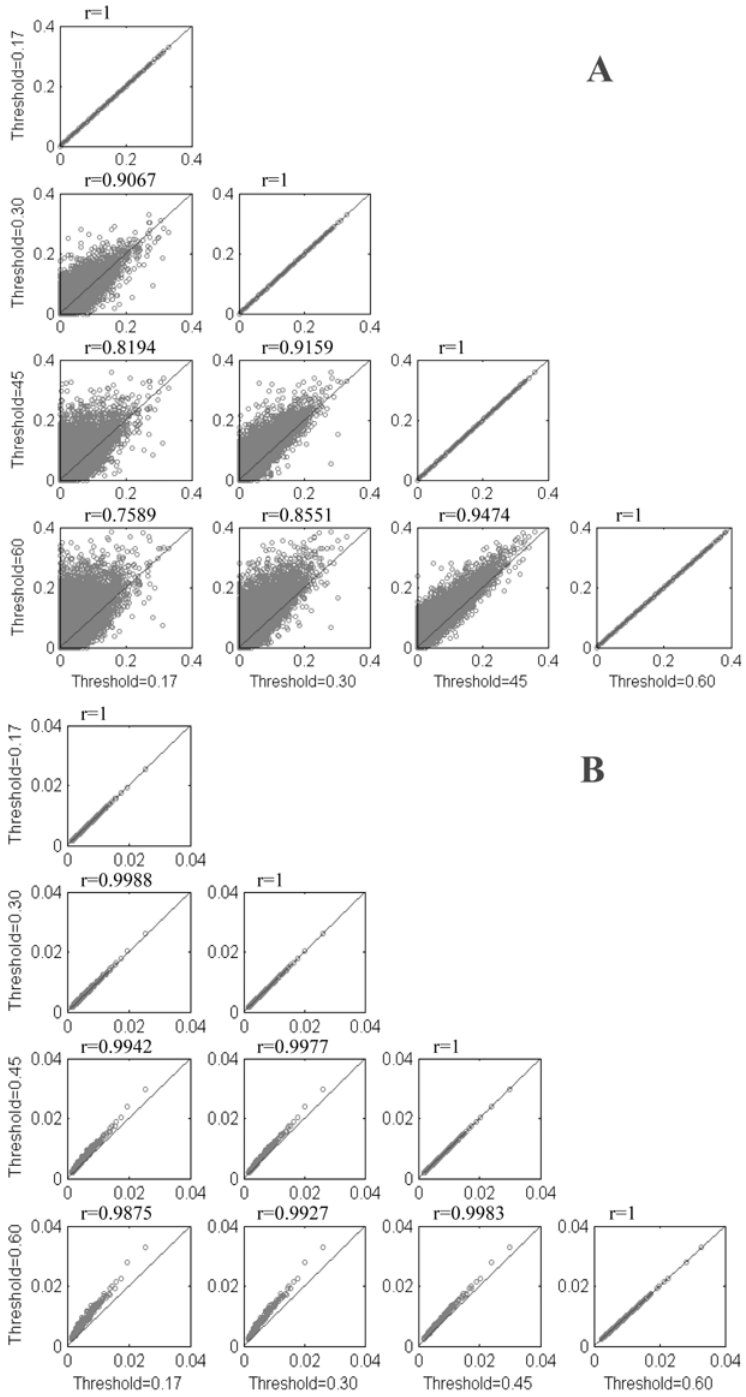

Figure 1. Scatter plots for comparing missing call rates between calling results with thresholds. The missing call rates calculated on individual SNPs (A) and samples (B) from genotype calling results of BRLMM with different thresholds for the Affy500K raw data of the 270 HapMap samples are plotted for pair-wise comparisons. The diagonal lines indicate that the missing call rates were the same in the two compared calling results. The Pearson correlation coefficients between the missing call rates of the two compared calling results are given on the top of corresponding scatter plots.

scatter plots. It can be seen from figure 1 that using different thresholds generated inconsistent genotype calls from the exact same raw data. *t*-tests were performed to determine whether the two sets of missing call rates from a normal distribution could have the same mean when the standard deviations are unknown but assumed equal. The resulting *P* values for the comparisons were less than 0.0001, indicating that missing call rates of individual samples and SNPs are statistically different. Furthermore, it was observed that the inconsistency (defined as 1-*r*) of missing call rates were positively related to the corresponding differences of thresholds, and negatively related to the sum of thresholds of the compared calling experiments, as shown in figure 2.

Comparing missing call rates can only assess the effect of threshold changes on missing calls. Since homozygote, heterozygote, and variant homozygote are possible results for a genotype call, we determined the effect of threshold on the ability to consistently call the genotypes. To evaluate the effect of threshold parameters on successful calls, concordance of calling results with different thresholds was analysed (table 1). The threshold affected successful genotype calls since the discordant calls existed for all of the comparisons. Moreover, the heterozygous genotype concordances were more affected than homozygous genotype concordances.

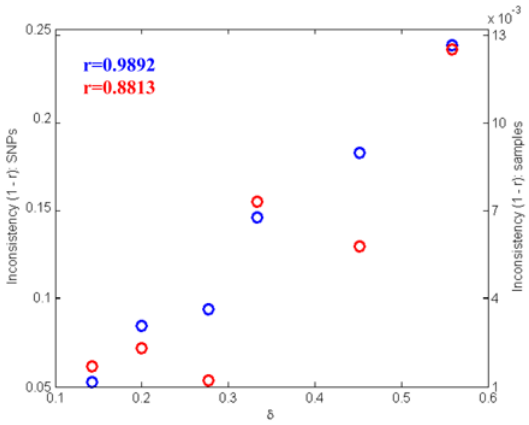

Figure 2. Relationship between missing call rate inconsistency and the difference of the thresholds used in BRLMM. Inconsistency (1-*r*) of the missing call rates calculated on individual SNPs (left y-axis and blue points) and samples (right y-axis and red points) from genotype calling results of BRLMM with different thresholds for the Affy500K raw data of the 270 HapMap samples were plotted against δ. The Pearson correlation coefficient, *r*, was calculated between the missing call rates of the two compared calling results *i* and *j*. The δ was defined and calculated as $\delta = \frac{Threshold^i - Threshold^j}{Threshold^i + Threshold^j}$.

Effect of changing the size parameter

Sizes of 5000, 10,000, 15,000 and 20,000 were used to estimate the prior distribution in BRLMM. The overall missing call rates obtained when using sizes of 5000, 10,000, 15,000, and 20,000 were 0.5164%, 0.5154%, 0.5170%, and 0.5163%, respectively. Thus, the number of SNPs used to estimate the prior distribution for BRLMM did affect the genotype calling result, but its influence was very small. However, as noted above, the overall missing call rates are not informative enough to assess their distribution in the data set. Again, the effect was compared using one-against-one comparisons of the distributions of missing call rates on individual samples and SNPs.

Table 1. Comparison of discordant successful calls.

Threshold	0.17	0.30	0.45	0.60
0.17	–	19369	51139	86704
0.30	41	–	22270	50325
0.45	81	40	–	16775
0.60	140	80	41	–

Above the diagonal, discordant hetrozygous calls; below the diagonal, discordant homozygous calls.

The comparisons of missing call rates of individual SNPs and samples between the results obtained by using sizes of 5000, 10,000, 15,000, and 20,000 are given in the scatter plots of figure 3,A&B, respectively. The Pearson correlation coefficients of corresponding comparisons were calculated and are shown on the top of the scatter plots. It can be seen from figure 3 that the same genotype calls were not obtained from the exact same raw data when using different sizes. However, the inconsistency of missing calls was much smaller compared to the effect of threshold (figure 3 versus figure 1). The missing call rates were very similar between calling results from using different sizes. The *P* values from *t*-tests for the comparisons were larger than 0.1, indicating that the differences in the missing call rates on both samples and SNPs were not statistically significant. To evaluate the effect of size on successful calls, the concordance between calling experiments with different sizes was analysed. It was observed that the successful calls were exactly same among different sizes (results not shown).

Quality of the raw data

The quality of the raw data is also important for comparative analyses and interpretation. The quality control (QC) scores of the 270 Nsp CEL files and of the 270 Sty chip CEL files of the 270 HapMap samples were calculated using the DM algorithm. The average QC scores for Nsp and Sty CEL files are 97.58 and 98.26, respectively. The lowest QC scores for Nsp and Sty CEL files are 93.49 and 93.18, respectively. The Affymetrix default QC cut-off score is 93. Therefore, we confirmed a high QC of the raw data and used all CEL files of 270 HapMap samples in our study.

Propagation of algorithmic parameter effect to associated markers

The objective of a GWAS is to identify genetic markers associated with a phenotype. It is critical to assess whether and how the algorithmic parameter's changes

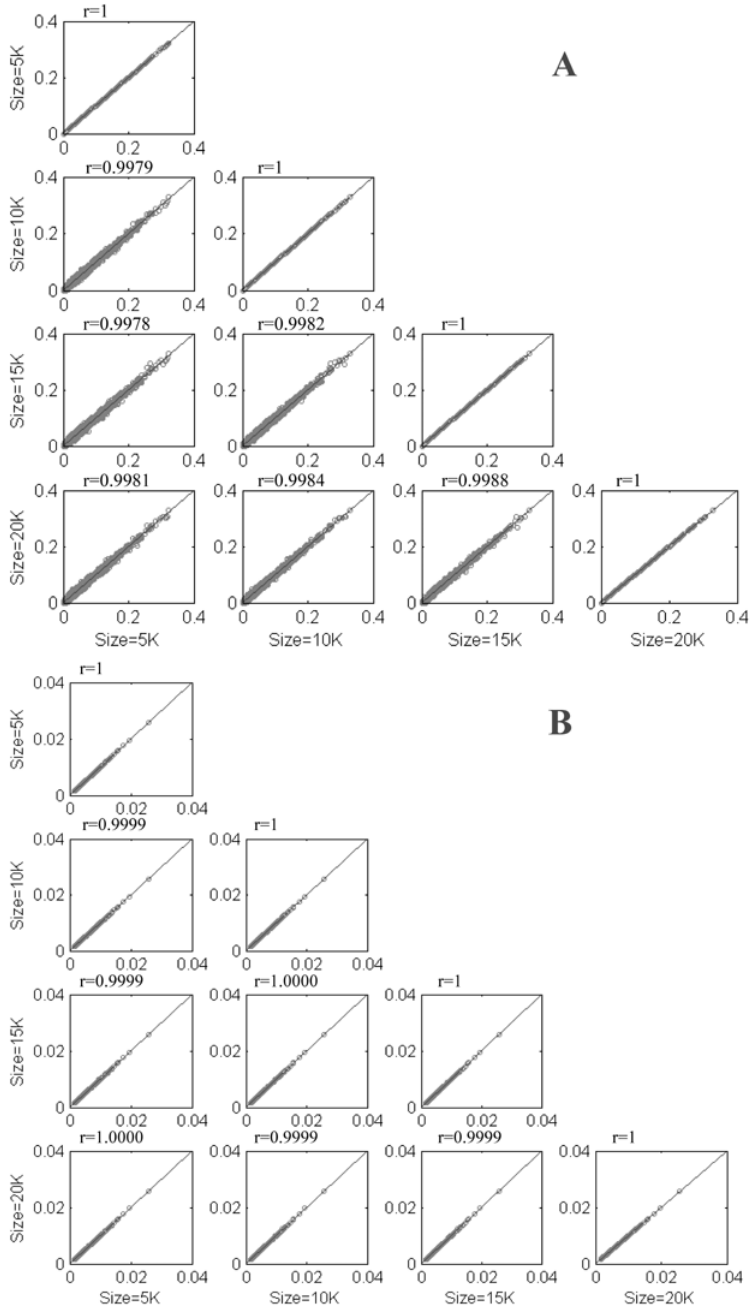

Figure 3. Scatter plots for comparing missing call rates between calling results with different sizes. The missing call rates calculated on individual SNPs (A) and samples (B) from genotype calling results of BRLMM with different sizes for the Affy500K raw data of the 270 HapMap samples were plotted for pairwise comparisons. The diagonal lines indicate that the missing call rates were the same in the two compared calling results. The Pearson correlation coefficients between the missing call rates of the two compared calling results are given on the top of corresponding scatter plots.

propagate to the significantly associated SNPs identified in the downstream analysis. Three case–control mimic association analyses were conducted for each of the calling results with different thresholds and sizes to assess the propagation of genotype inconsistency to the significantly associated SNPs. After removal of low-quality SNPs by quality control assessment, each of the three population groups (European, Asian and African) was set as 'case' while the

other two groups were set as 'control'. Association analyses were conducted to identify SNPs that can differentiate the 'case' population from the control population. Different lists of SNPs associated with the same population group, identified using the genotype calling results with different thresholds and sizes, were compared using Venn diagrams.

The comparison of the significantly associated SNPs obtained from calling results with different thresholds are given in figure 4. It is clear that threshold effect on genotype calling propagated into the downstream association analyses, since for any statistical tests (genotypic, left column; allelic, right column) and case–control mimics (Asian as case, first row; European as case, second row; African as case, last row) there was a discordance of significantly associated SNPs identified between different thresholds. Moreover, the frequency of discordant SNPs between different thresholds was positively related to the difference of thresholds used in BRLMM, as shown in figure 5.

The comparisons of the significantly associated SNPs obtained from calling results with different sizes are shown in figure 6. It can be seen that the discordant significantly associated SNPs between different sizes were 0.082–0.113%, much less compared to the corresponding ones for the threshold effect (1.004%–2.764%).

Discussion

GWAS is increasingly used to identify loci containing genetic variants associated with common diseases and drug responses. The number of SNPs interrogated in a GWAS has grown from thousands to millions; for example, the newest Affymetrix SNPs array 6.0 contains ~2 million probe sets. At the same time, the allele frequency difference of disease-associated or drug-associated SNPs is usually very small. Therefore, a very small error rate introduced in genotypes by genotype calling algorithms may result in inflated type I and type II errors in the downstream association analysis.

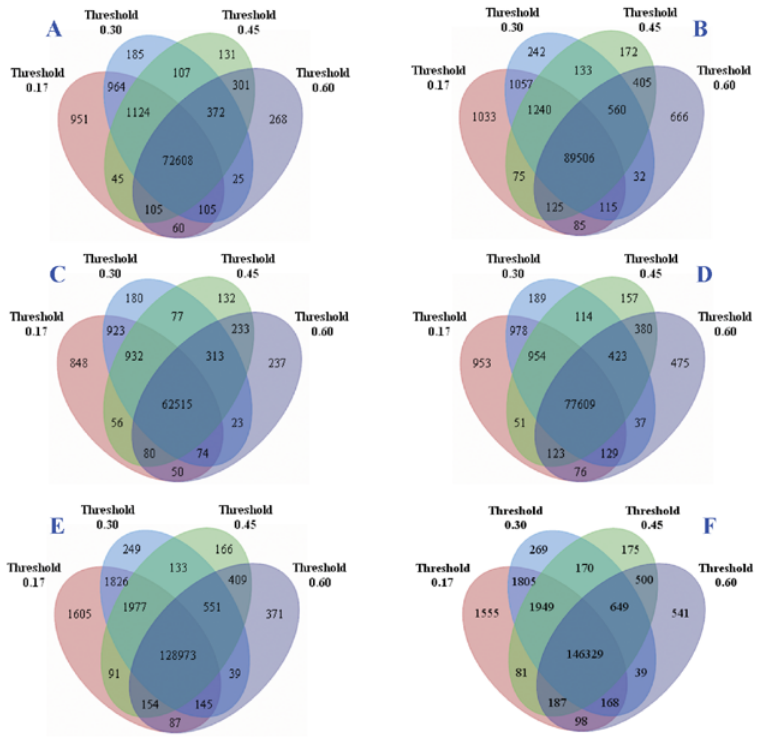

Figure 4. Venn diagrams for comparisons of the significantly associated SNPs identified in the association analyses using the genotype calling results with thresholds. The numbers in ellipses are the significantly associated SNPs identified in association analyses using calling results from different thresholds: pink for threshold = 0.17, blue for threshold = 0.30, green for threshold = 0.45, and purple for threshold = 0.60. Numbers in the sections of ellipses represent the significantly associated SNPs shared by the corresponding thresholds. Left column (A, C, E) are the results from genotypic associations; right column (B, D, F) are from the allelic associations. First row (A, B), the association analyses results using Asian population as case; second row (C, D), the results using European as case; and the last row (E, F), the results using African as case.

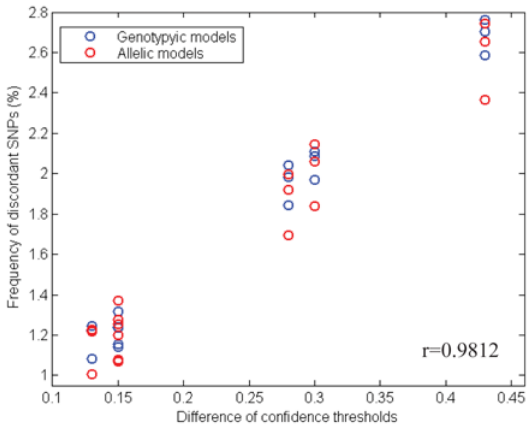

Figure 5. Relationship between discordant significantly associated SNPs and the difference of the thresholds used in BRLMM. Frequency of the discordant significantly associated SNPs identified by genotypic association (blue points) and allelic association (red points) from genotype calling results of BRLMM for the Affy500K raw data of the 270 HapMap samples with two different thresholds i and j, defined and calculated as $\frac{100(n_i^j+n_j^i)}{(n_i+n_j)}$ where n_i and n_j are numbers of significantly associated SNPs identified from results with thresholds i and j respectively, n_i^j is number of SNPs significant from threshold i but not significant from threshold j, and n_j^i is number of SNPs significant from threshold j but not significant from threshold i, were plotted against the difference of thresholds ($threshold^i - threshold^j$). Given at the bottom-right corner is the Pearson correlation coefficient.

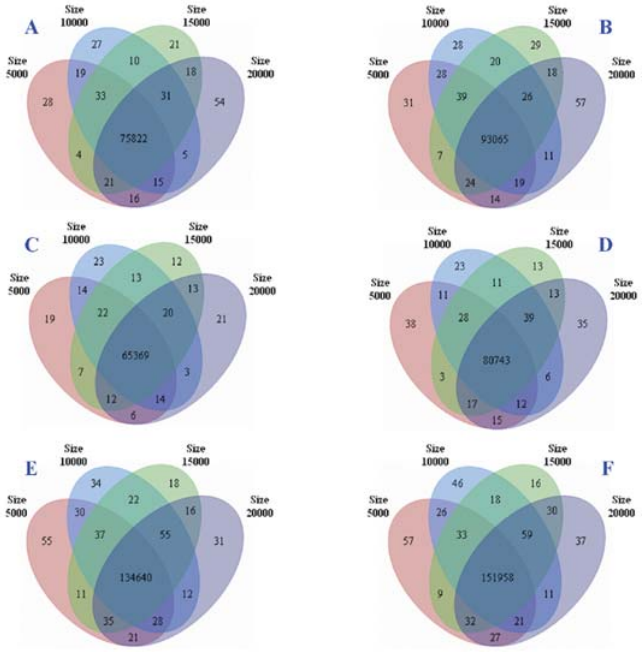

Figure 6. Venn diagrams for comparisons of the significantly associated SNPs identified in the association analyses using the genotype calling results with different sizes. The numbers in ellipses are the significantly associated SNPs identified in association analyses using calling results from different sizes: pink for size = 5,000, blue for size = 10,000, green for size = 15,000, and purple for size = 20,000. Numbers in the sections of ellipses represent the significantly associated SNPs shared by the corresponding sizes. Left column (A, C, E) are the results from genotypic associations; right column (B, D, F) are from the allelic associations. First row (A, B), the association analyses results using Asian population as case; second row (C, D), the results using European as case; and the last row (E, F), are the results using African as case.

It is important to know the robustness of a genotype calling algorithm. Availability of different parameters in the BRLMM genotype calling algorithm makes it vital to be aware of the genotype inconsistency caused by using altered parameter values and their effect on GWAS results.

A heterozygous genotype carries a rare allele. Therefore, the robustness of a heterozygous calling reduces false positive associations. Our studies revealed that heterozygous genotype calling was sensitive to algorithmic parameters and thus algorithms that reduce algorithmic parameters effects and maintain high call rates and accuracy are needed. Previously, we analysed both the batch size and the batch composition affect the genotype calling results in GWAS using BRLMM (Hong *et al*. 2008). Batch size and batch composition effects were found to be a more severe problem on samples and SNPs with lower call rates, and on heterozygous genotype calls. Therefore, the influences of batch effects of Affy500K arrays are correlated with the influence of genotyping errors.

Genotype discordance was found in both missing calls and successful calls. Our study showed the propagation of discordant genotypes to the significantly associated SNPs was caused by both sources of discordance. Our observations suggest that there is a room for improvements on both call rate and accuracy of calling algorithms.

An interesting observation in our study was that more significantly associated SNPs were identified in the model using African as case (figures 4 and 6). In the HapMap samples it is well known that the Yoruban is more genetically distinct than the Asian and European. However, discordant rates of the significantly associated SNPs for the African model were lower than the Asian and European models (figure 7). Therefore, discordance in genotypes might be amplified more in the significantly associated SNPs for weaker traits than for stronger traits. Comparing with the population differences of the HapMap samples used in our study, traits of current GWAS are usually much weaker, and a smaller number of concordant significantly associated SNPs are expected.

Besides genotyping errors, another factor to be considered in genotype calling for Affy500K arrays is missing call bias (MCB). Because MCB often leads to biased conclusions in the subsequent analyses, including estimation of allele/genotype frequencies, the measurement of Hardy–Weinberg equilibrium and association tests under various modes of inheritance relationships, MCB usually leads to power loss in association tests (Fu *et al*. 2009). In addition, unaccounted sample failure as well as hidden population structure can introduce misleading signals that may mimic

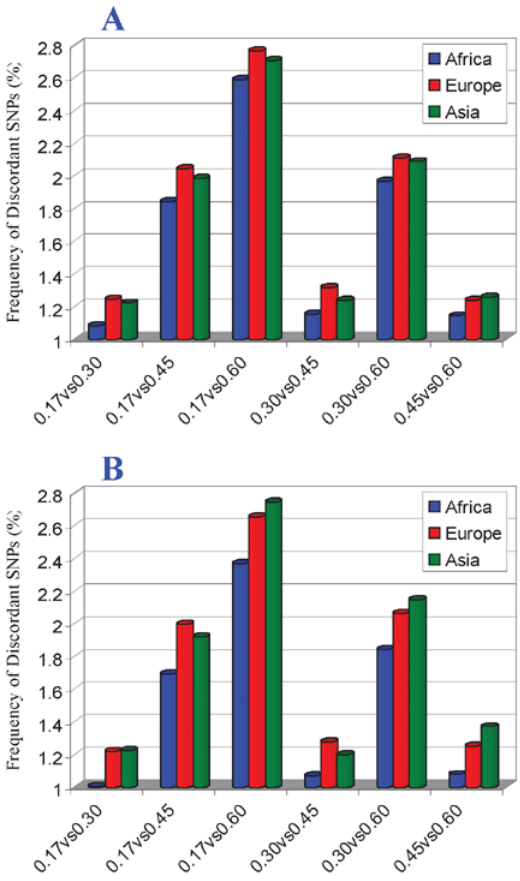

Figure 7. Frequency of discordant significantly associated SNPs identified by the association analysis for the pair-wise comparisons between different thresholds used in BRLMM. Frequency of discordant significantly associated SNPs identified by genotypic association (A) and allelic association (B) from genotype calling results of BRLMM for the Affy500K raw data of the 270 HapMap samples between two different thresholds *i* and *j* for the pair-wise comparisons indicated at the x-axis for the three case-population-based models (blue, African; red, European; green, Asian) are drawn in bars.

genuine association (Teo 2008). Given the major challenge of separating the many false positive from the few true positive associations with the disease phenotype in GWAS, an important strategy has been replication of results in independent samples (Chanock *et al*. 2007).

As demonstrated above, there are threshold effects when using the BRLMM algorithm that alters genotype calling results of GWAS. The larger the difference between the thresholds used, the greater the effect; the more confidence of DM calls used (smaller threshold), the more consistent the genotypes. However, the size used to estimate the prior distribution in BRLMM does not have a statistically significant effect on the genotype calling results and the significantly associated SNPs identified in the downstream association analysis. Therefore, our study suggests that smaller and consistent thresholds should be used to make genotype calls for GWAS using data from the Affy500K coupled with the BRLMM algorithm.

Acknowledgements

We thank Drs Jim Kaput, James Fuscoe, James Chen, Tao Han, Joshua Xu of NCTR/FDA; Federico Goodsaid, Isaam Zineh, Sue Jane Wang; and Li Zhang of CDER/FDA; Kelci Miclaus, Wendy Czika and Russell D. Wolfing of SAS Institute; Silvia C. Vega of Rosetta BioSoftware; Marco Chieric of Fondazione Bruno Kessler; Christophe G. Lambert of Golden Helix; Ansar Jawaid of AstraZeneca; Uwe Scherf, Lakshmi Vishnuvajjala, Arkendra De and Lakshman Ramamurthy of CDRH/FDA; and Keith Nangle, Meg E. Ehm, and Gbenga R. Kazeem of GlaxoSmithKline for fruitful discussions.

References

Affymetrix 2006 *BRLMM: an improved genotype calling method for the GeneChip® Human Mapping 500K array set*, Revision version 1.0. April 14, 2006. URL: http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf

Arking D. E., Cutler D. J., Brune C. W., Teslovich T. M., West K., Ikeda M. *et al*. 2008 A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am. J. Hum. Genet*. **82**, 160–164.

Benjamini Y. and Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B*. **57**, 289–300.

Buch S., Schafmayer C., Völzke H., Becker C., Franke A., von Eller-Eberstein H. *et al*. 2007 A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nature Genet*. **39**, 995–999.

Butcher L. M., Davis O. S., Craig I. W. and Plomin R. 2008 Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes Brain Behav*. **7**, 435–446.

Cargill M., Schrodi S. J., Chang M., Garcia V. E., Brandon R., Callis K. P. *et al*. 2007 A large-scale genetic association study confirms *IL12B* and leads to the identification of *IL23R* as psoriasis-risk genes. *Am. J. Hum. Genet*. **80**, 273–290.

Chanock S. J., Manolio T., Boehnke M., Boerwinkle E., Hunter D. J., Thomas G. *et al*. (NCI-NHGRI working group on replication in association studies) 2007 Replicating genotype-phenotype associations. *Nature* **447**, 655–660.

Di X., Matsuzaki H., Webster T. A., Hubbell E., Liu G., Dong S. *et al*. 2005 Dynamic model based algorithms for screening and genotyping over 100K SNPs on oligonucleotide microarrays. *Bioinformatics* **21**, 1958–1963.

Duerr R. H., Taylor K.D., Brant S. R., Rioux J. D., Silverberg M. S., Daly M. J. *et al*. 2006 A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science* **314**, 1461–1463.

Easton D. F., Pooley K. A., Dunning A. M., Pharoah P. D., Thompson D., Ballinger D. G. *et al*. 2007 Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093.

Frayling T. M., Timpson N. J., Weedon M. N., Zeggini E., Freathy R. M., Lindgren C. M. *et al*. 2007 A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894.

Fu W., Wang Y., Wang Y., Li R., Lin R. and Jin L. 2009 Missing call bias in high-throughput genotyping. *BMC Genomics* **10**, 106.

Gold B., Kirchhoff T., Stefanov S., Lautenberger J., Viale A., Garber J. *et al*. 2008 A genome-wide association study provides evidence for a breast cancer risk at 6q22.33. *Proc. Natl. Acad. Sci. USA* **105**, 4340–4345.

Grupe A., Abraham R,, Li Y., Rowland C., Hollingworth P., Morgan A. *et al*. 2007 Evidence for novel susceptibility genes for late-onset Alzheimers disease from a genome-wide association study of putative functional variants. *Hum. Mol. Genet*. **16**, 865–873.

Gudmundsson J., Sulem P., Manolescu A., Amundadottir L. T., Gudbjartsson D., Helgason A. *et al*. 2007 Genome-wide association study identifies a second breast cancer susceptibility variant at 8q24. *Nature Genet*. **39**, 631–637.

Hampe J., Franke A., Rosenstiel P., Till A., Teuber M., Huse K. *et al*. 2007 A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature Genet*. **39**, 207–211.

Hong H., Su Z., Ge W., Shi L., Perkins R., Fang H. *et al*. 2008 Assessing batch effects of genotype calling algorithm BRLMM for the Affymetrix GeneChip human mapping 500K Array Set using 270 HapMap samples. *BMC Bioinformatics* **9**, S17.

Hunter D. J., Kraft P., Jacobs K. B., Cox D. G., Yeager M., Hankinson S. E. *et al*. 2007 Genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nature Genet*. **39**, 870–874.

Kayser M., Liu F., Janssens A. C., Rivadeneira F., Lao O., van Duijn K. *et al*. 2008 Three genome-wide association studies and a linkage analysis identify *HERC2* as a human iris color gene. *Am. J. Hum. Genet*. **82**, 411–423.

Klein R. J., Zeiss C., Chew E. Y., Tsai J. Y., Sackler R. S., Haynes C. *et al*. 2005 Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389.

Moore A. F., Jablonski K. A., McAteer J. B., Saxena R., Pollin T. I., Franks P. W. *et al*. 2008 Extension of type 2 diabetes genome-wide association scan results in the Diabetes Prevention Program. *Diabetes* **57**, 2503–2510.

Moskvina V., Craddock N., Holmans P., Owen M. J. and O'Donovan M. C. 2006 Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum. Hered*. **61**, 55–64.

Raelson J. V., Little R. D., Ruether A., Fournier H., Paquin B., Eerdewegh P. V. *et al*. 2007 Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proc. Natl. Acad. Sci. USA* **104**, 14747–14752.

Rioux J. D., Xavier R. J., Taylor K. D., Silverberg M. S., Goyette P., Huett A. *et al*. 2007 Genome-wide association study identifies new susceptibility loci for Crohn's disease and implicates autophagy in disease pathogenesis. *Nature Genet*. **39**, 596–604.

Saxena R., Voight B. F., Lyssenko V., Burtt N. P., de Bakker P. I., Chen H. *et al*. 2007 Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride level. *Science* **316**, 1331–1336.

Scott L., Mohlke K. L., Bonnycastle L. L., Willer C. J., Li Y., Duren W. L. *et al*. 2007 A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345.

Sladek R., Rocheleau G., Rung J., Dina C., Shen L., Serre D. *et al*. 2007 A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885.

Smyth D. J., Cooper J. D., Bailey R., Field S., Burren O., Smink L. J. *et al*. 2006 A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nature Genet*. **38**, 617–619.

Steinthorsdottir V., Thorleifsson G., Reynisdottir I., Benediktsson R., Jonsdottir T., Walters G. B. *et al*. 2007 A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nature Genet*. **39**, 770–775.

Teo Y. Y. 2008 Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Curr. Opin. Lipidol*. **19**, 133–143.

The International HapMap Consortium 2005 A haplotype map of the human genome. *Nature* **437**, 1299–1320.

The International HapMap Consortium 2007 A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–862.

Todd A. J., Walker N. M., Cooper J. D., Smyth D. J., Downes K., Plagnol V. *et al*. 2007 Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nature Genet*. **39**, 857–864.

Tomlinson I., Webb E., Carvajal-Carmona L., Broderick P., Kemp Z., Spain S. *et al*. 2007 A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nature Genet*. **39**, 984–988.

Uda M., Galanello R., Sanna S., Lettre G., Sankaran V. G., Chen W. *et al*. 2008 Genome-wide association study shows *BCL11A* associated with persistent fetal hemoglobin and amelioration of the phenotype of β-thalassemia. *Proc. Natl. Acad. Sci. USA* **105**, 1620–1625.

van Heel D. A., Franke L., Hunt K. A., Gwilliam R., Zhernakova A., Inouye M. *et al*. 2007 A genome-wide association study for celiac disease identifies risk variants in the region harbouring *IL2* and *IL21*. *Nature Genet*. **39**, 827–829.

Wellcome Trust Case Control Consortium 2007 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678.

Winkelmann J., Schormair B., Lichtner P., Ripke S., Xiong L., Jalilzadeh S. *et al*. 2007 Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nature Genet*. **39**, 1000–1006.

Yang H. H., Hu N., Taylor P. R. and Lee M. P. 2008 Whole genome-wide association study using Affymetrix snp chip: a two-stage sequential selection method to identify genes that increase the risk of developing complex diseases. *Clin. Bioinform*. **141**, 23–35.

Yeager M., Orr N., Hayes R. B., Jacobs K. B., Kraft P., Wacholder S. *et al*. 2007 Genome-wide association study of breast cancer identifies a second risk locus at 8q24. *Nature Genet*. **39**, 645–649.

Zanke B. W., Greenwood C. M., Rangrej J., Kustra R., Tenesa A., Farrington S. M. *et al*. 2007 Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nature Genet*. **39**, 989–994.

Zeggini E., Weedon M. N., Lindgren C. M., Frayling T. M., Elliott K. S., Lango H. *et al*. 2007 Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341.

Received 5 August 2009; accepted 26 November 2009

Published on the Web: 1 April 2010