Published in final edited form as:

Stat Biosci. 2009 November 1; 1(2): 214–227. doi:10.1007/s12561-009-9011-4.

The null distributions of test statistics in genomewide association studies

Xiang Chen and Heping Zhang

Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520-8034, U.S.A

SUMMARY

Genomewide association (GWA) studies assay hundreds of thousands of single nucleotide polymorphisms (SNPs) simultaneously across the entire genome and associate them with diseases, other biological or clinical traits. The association analysis usually tests each SNP as an independent entity and ignores the biological information such as linkage disequilibrium. Although the Bonferroni correction and other approaches have been proposed to address the issue of multiple comparisons as a result of testing many SNPs, there is a lack of understanding of the distribution of an association test statistic when an entire genome is considered together. In other words, there are extensive efforts in hypothesis testing, and almost no attempt in estimating the density under the null hypothesis. By estimating the true null distribution, we can apply the result directly to hypothesis testing; better assess the existing approaches of multiple comparisons; and evaluate the impact of linkage disequilibrium on the GWA studies. To this end, we estimate the empirical null distribution of an association test statistic in GWA studies using simulated population data. We further propose a convenient and accurate method based on adaptive spline to estimate the empirical value in GWA studies and validate our findings using a real data set. Our method enables us to fully characterize the null distribution of an association test that not only can be used to test the null hypothesis of no association, but also provide important information about the impact of density of the genetic markers on the significance of the tests. Our method does not require users to perform computationally intensive permutations, and hence provides a timely solution to an important and difficult problem in GWA studies.

Keywords

critical value; generalized extreme-value distribution; genomewide association

1. Introduction

The characterization of the human genome and advances in high throughput technology for genetic variations enable us to perform genomewide scans for genes underlying common diseases. To perform those scans, deoxyribonucleic acid (DNA) sequence data are collected for study subjects. The sequence consists of chemical bases among four possible choices: adenine (A), guanine (G), cytosine (C), and thymine (T). Human genome has about 3 billion base pairs. More than 99 percent of the bases are the same in all people, and those less than 1 percent of the varying bases are single nucleotide polymorphisms (SNPs).

Since the first attempt of genomewide association (GWA) study in 2002(Ozaki et al., 2002), there has been a dramatic increase in the number of both the launched projects and the identified loci through GWA approaches. Successes of GWA studies include the confirmatory identification of the Complement Factor H in age-related macular degeneration(Edwards et al., 2005; Haines et al., 2005; Klein et al., 2005), myocardial infarction(Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007; The Wellcome Trust Case Control Consortium, 2007), and type 2 diabetes(Saxena et al., 2007; Scott et al., 2007; Sladek et al., 2007; Zeggini et al., 2007). A more recent study found that the same sequence variant on 9p21 is associated with abdominal aortic aneurysm and intracranial aneurysm(Helgadottir et al., 2008).

GWA studies assay hundreds of thousands of SNPs simultaneously across the entire genome in relation to disease or other biological traits. As a result, hundreds of thousands of hypotheses are simultaneously tested. Both planning and execution of GWA studies depend on the distribution of the association test under the null hypothesis in order to determine the sample size in the study design and the p-value in hypothesis testing. The efforts to deal with the association tests in GWA studies have been mostly mechanistic such as the correction of the significant level. We are aimed to characterize the null distribution of the test statistic by incorporating the information from the entire genome. For example, our result provides a benchmark for the performance of the mechanistic methods when the biological information is integrated.

Although various approaches have been proposed to deal with the multiple comparison issue (Gao et al., 2009; Moskvina and Schmidt, 2008; Sabatti, Service and Freimer, 2003; Storey and Tibshirani, 2003; The Wellcome Trust Case Control Consortium, 2007; Wacholder et al., 2004), the traditional Bonferroni correction is still the most commonly used correction method in GWA studies(Cauchy et al., 2008; Ke, Taylor and Cardon, 2008; Klein et al., 2005; Stranger et al., 2005), particularly in the computation of sample size. With the Bonferroni correction, the conventional significance level (α level) is divided by the total number of tests performed (i.e., the number of SNPs). It ensures that the possibility of false identification of association due to chance alone is less than the pre-specified level α , regardless of the dependency among multiple tests. In other words, the Bonferroni correction ignores the information in specific data. For example, SNPs in many regions of the genome with a high density are known to be correlated due to linkage disequilibrium (LD), because the SNPs that are physically close tend to be inherited together, instead of independently. Consequently the Bonferroni correction tends to be increasingly conservative with higher SNP density and will reduce power because only the relatively strong associations could be declared statistically significant(Nakagawa, 2004). Thus, the power of GWA studies can be improved if the LD between SNPs can be taken into account.

Instead of adjusting the comparison-wise p-value or permuting data for every GWA data set, we construct the distribution of the maximum absolute value (maximum statistic) of a specific statistic (such as allelic χ^2 statistic, genotype χ^2 statistic, and Z_T statistic from the Armitage trend test) across all SNPs in common platforms for GWA studies under the null hypothesis (no association between any SNP and the phenotype). The maximum statistic captures information regarding the number of SNPs as well as the LD between SNPs in GWA studies. Intuitively, the maximum statistic reflects how large a test statistic can be across the genome under the null hypothesis. Using a variant of the maximum statistic, Dudbridge and Gusnanto estimated the genomewide significance threshold (p-value) of 7.2×10^{-8} for achieving 5% FWER in UK Caucasian population(Dudbridge and Gusnanto, 2008). Conneely and Boehnke developed a method for addressing the correlated nature of SNPs using the maximum statistic (Conneely and Boehnke, 2007).

A good approximation of the null distribution is the generalized extreme-value (GEV) distribution (Jenkinson, 1955) since we used genomewide maximum of a specific statistic to generate the null distributions. Given the complexity of the genome, it may not be feasible to derive the parameters and quantiles (i.e., critical values) of the distribution for any specific maximum statistic. When data are already collected, the most commonly used approach to estimating the critical value is permutation, which yields an empirical null distribution (Chen et al., 2007; Churchill and Doerge, 1994). However, permutation based approach has several shortcomings: (a) it is computationally intensive, even with the latest computing technology; and (b) when it generates a specific critical value for a given significance level, unlike the critical value from a theoretical distribution (e.g., the standard normal distribution), the critical value from the permutation procedure may be an undesirable degree of variation. Hence, the p-value obtained from the permutation is subject to uncertainty, which has not received the necessary attention. In fact, as demonstrated in Figure 3, the extent of this uncertainty is great enough to change the significance of a finding. Thus, even if we use the permutation, we need an empirical critical value that has a much lower variability.

Therefore, our goal is to characterize the null distribution of the test statistic in GWA studies, which in turn can be used to derive genomewide critical values that are easy to use (the advantage of the Bonferroni correction), and that reflect the genomic data (the advantage of the permutation), and that are accurate and has a low level of uncertainty. For these motivations, instead of performing the dataset specific permutations, we take advantage of the data from the HapMap project (The International HapMap Consortium, 2003), recent advances in genomewide SNP data simulations (Durrant et al., 2004; Li and Li, 2008), and high throughput computing facility. Those resources enable us to generate population data with arbitrary number of subjects and arbitrary number of SNPs (up to the number of SNPs included in the HapMap project) under the null hypothesis. This is possible because of an intuitive, yet critical observation: under the null hypothesis, the trait and the genotypes are not associated. Thus, the null distribution of any association test statistic is not trait specific, and it depends on the case-control ratio only. In the current report, we estimate the empirical null distribution for several most useful test statistics in GWA studies, namely allelic χ^2 test, genotype χ^2 test, and the Armitage trend tests for dominant, additive and recessive models.

After studying the relationship between the number of SNPs sampled across the genome and the empirical null distributions, we further proposed a model that estimates the null distribution (and hence any empirical critical value) of a test statistic for GWA studies, which has small bias and variation, and does not require the users to perform computationally intensive permutations.

2. Material and Methods

2.1 Phased input data

All phased data were obtained from the international HapMap project (http://ftp.hapmap.org/phasing/2007-08_rel22/phased/, Release 22, on NCBI B36 assembly). The HapMap project has obtained detailed genotype and haplotype information from a limited number of subjects in several different populations (90 subjects in the JPT+CHB population, 30 trios in the CEU population and 30 trios in the YRI population) (The International HapMap Consortium, 2003). All SNPs on autosomes (Chromosomes 1–22) were included in this study. For both the CEU (2,543,909 SNPs) and the YRI (2,852,206 SNPs) populations, phased data from 60 unrelated individuals (parents of the 30 trios) were used. For the JPT+CHB (2,416,663 SNPs) population, phased data from 90 unrelated individuals were used.

2.2 Simulation of population samples

Association mapping relies on LD between disease loci and marker loci. Therefore, in the current study, it is critical to estimate the empirical null distribution using simulated data that have similar LD structure as seen in real data. Two popular algorithms aiming to maintain LD patterns in simulated data are the coalescent approach (Donnelly and Tavaré, 1995) and the moving-window approach (Durrant et al., 2004). Coalescent algorithms simulate genomic data from a theoretical population. However, the programs based on this approach are generally inefficient and time consuming. The moving-window algorithms generate simulated genotypes from phased data as input. A recent implementation of a moving-window algorithm suggested that it is more efficient than the coalescent approach and the simulated data faithfully follow the local LD structure of the input data (Li and Li, 2008).

The population sample simulations were implemented using the algorithms described in (Durrant et al., 2004; Li and Li, 2008) with a window size of 5. In each simulation, different numbers of subjects from a specified population (1000, 2000, and 4000 for balanced samples, and 1500 and 3000 for unbalanced samples) were generated.

2.3 Sampling of SNPs for sub-data set

From the simulated full data set, we generated a series of subsets of data with the same number of subjects but fewer SNPs (with 1/2, 1/4, 1/8, 1/16, 1/32, 1/64 and 1/128 number of SNPs) by sampling one SNP in every 2, 4, 8, 16, 32, 64 and 128 consecutive SNPs. This allows us to evaluate the impact of the number of SNPs on the critical values.

2.4 Generation of an empirical null distribution

For each balanced simulated data, a random half of the samples were assigned to the case group and the rest half were assigned to the control group (permutation). For each unbalanced simulated data, one third of the samples were randomly assigned to the case group and the rest two thirds were assigned to the control group. Allelic association tests, genotype association tests and Armitage trend tests (with dominant, additive and recessive models) were performed on each individual SNP, and the maximum test statistic from all SNPs in the dataset was used to generate the empirical null distribution of a specific association test for a different dataset. One thousand permutations were carried out to construct an empirical null distribution and to estimate the empirical critical value. We independently generated 100 null distributions for each population.

2.5 GEV modeling

The generalized extreme value distribution (GEV) is the limit distribution of maxima of a sequence of independent, identically distributed random variables. We used GEV to model the null distribution of the maximum statistic (for both the genotype χ^2 test and Armitage trend tests). For a specific experimental design, we used the `fgev` procedure in the R `evd` package to estimate the parameters of the corresponding GEV distribution through maximum-likelihood fitting.

2.6 Nonparametric estimate of a curve

In order to model the relationship between the empirical critical value and the number of the genotyped markers, we used a non-parametric method to fit a curve to the observed data points that are correlated. The method is referred to as “multivariate adaptive splines for analysis of longitudinal data” (MASAL) (Zhang, 1999).

In the current context, the goal of MASAL is to find the most efficient (i.e., a simple and most accurate form) piecewise linear approximation to an underlying smooth curve. The idea is

similar to the numerical integration. The major difference between MASAL and other curve smoothing methods is that MASAL allows the users to introduce a correlation structure. In this study, we used the correlation structure $e^{c_0+c_1 \times \log_{10}(\#SNP)}$, where c_0 and c_1 are estimated in MASAL. Briefly, let y denote the response variable, which is the critical values at different SNP densities, and x be the logarithms of the number of SNPs. We consider a nonparametric regression model

$$y_{ix}=f(x)+e_{ix},$$

where y_{ix} is the critical value when e^x SNPs are typed in the i th simulation, f is an unknown smooth function and e_{ix} is an error term with mean zero with the assumed correlation structure. MASAL estimates the function f from the following class through stepwise procedures:

$\beta_0 + \sum_{v=1}^V \beta_v B_v(x)$, where β_v is the regression coefficient, and $B_v(x)$ is the basis function defined as either one of the following functions or their product, $(x-a)^+$ and x . Here, $(x-a)^+ = \max(0, x-a)$. V is the total number of bases that also needs to be estimated from the data.

2.6 Estimation of the empirical critical value for the Cancer Genetic Markers of Susceptibility breast (CGEMS) cancer dataset

The CGEMS breast cancer dataset was obtained from <http://cgems.cancer.gov/>. This dataset contains 1231 case subjects and 1201 control subjects, who were genotyped with 541,327 autosomal SNPs using the Illumina HumanHap550 array. After removal of SNPs with excessive missing rates ($> 10\%$), a total of 532,075 SNPs were used to generate the empirical null distribution of the critical values. A thousand permutations were carried out to construct an empirical null distribution and to estimate the empirical critical value. The average and SD of the empirical critical values were estimated from 100 replicates.

3. Results

3.1 Empirical null distributions

Figure 1 displays a variety of empirical null distributions generated through genomewide simulations, reflecting design considerations in typical GWA studies. As expected, the shapes of all empirical null distributions are similar and all are of a bell shape with a heavy right tail, regardless of the test statistics (Panel A), the population (Panel B), the number of SNPs genotyped (Panel C), the number of samples (Panel D), or the case-control ratios (Panel E).. It is this consistency that enables us to establish an empirical null distribution for investigators to use when they design their GWA studies. Furthermore, we fitted the empirical distributions using GEV distribution and Figure 2 shows a Q-Q plot and a density plot between the modeled GEV distribution (parameters estimated by maximum likelihood) and its corresponding empirical distribution. It suggests that the GEV distribution is a good approximation for the empirical distribution.

3.2 Relationship between the experimental designs and the empirical null distribution

Figure 1B shows that for a specific statistical test, the null distribution curve for YRI population is shifted to the right compared to those for CEU and JPT+CHB populations. This trend is observed when the genome is densely genotyped (with more than 200,000 SNPs). This result is consistent with the observations that “LD extends to a similar and long extent in Asian as well as European samples” while the African-American samples showed lower LD (Gabriel et al., 2002). Therefore, our proposed method appropriately reflects and takes advantage of the

LD information. Due to the striking similarity among null distribution curves using different test statistics in different populations, we illustrate the relationship between other experimental designs (SNP density, sample size and case/control ratio) using the genotype χ^2 statistic in the JPT+CHB population in Panels C–E of Figure 1. Panel C indicates that the null distribution curve with more SNPs genotyped is shifted to the right compared to its counterpart with less SNPs genotyped. This general trend is expected, but our precise characterization is unique. Panel D reveals that the null distribution curve from a larger sample size is slightly shifted to the right compared to its counterpart from a smaller sample size. In addition, the empirical null distribution curve for the unbalanced experimental design is also slightly right shifted compared to the balanced experimental design (Panel E). The analyses using the YRI and CEU populations and other statistic yield similar trends (data not shown).

3.3 Estimation of empirical critical values

When designing a GWA study, we typically target at a genomewide significance level of 0.05 ($\alpha = 0.05$). Thus, it is useful to know how the number of SNPs that are relatively evenly spaced in the whole genome impacts the critical value to reach the targeted significance level and how conservative the Bonferroni correction is. Figure 3 displays a scatter plot where the x-axis is the number of SNPs from the three populations in the HapMap project [with 2000 simulated samples (1000 cases and 1000 controls)] and the y-axis is the empirical critical value for the genotype χ^2 statistic ensuring a family-wide type I error (FWER) of 0.05. We used a non-parametric method (Zhang, 1999) to draw the fitted curves that are also displayed in Figure 3. The dashed line represents the reference critical values if the Bonferroni correction is applied.

Two major observations are noteworthy. First, the Bonferroni corrected critical values are always and distinctly on the top of the fitted, empirical critical values, confirming that the Bonferroni method is conservative. The gap between the Bonferroni corrected critical values and our proposed critical values is relatively stable for less than 500,000 SNPs, and it accelerates with more than 500,000 SNPs. Given that most of the new SNP platforms have more than 500,000 SNPs, this is important to know. These gaps are due to basal correlations among SNPs in all populations, even when SNPs are sparsely sampled in the genome. Second, the fitted curves in all populations are essentially identical up to 200,000 SNPs, and then they diverge. In particular, the largest empirical χ^2 critical values are observed in the YRI population. This result is consistent with the observations using the null distribution curves. This importance of this comparison is not to point out the known fact that the Bonferroni correction is conservative, but to gain insights into the magnitude of the conservativeness when the SNP density varies.

The fitted curves for the critical values from each of the populations can serve as the empirical basis for the significance testing according to the number of genomewide SNPs in GWA studies. The SNPs used to construct Figure 3 are selected “evenly”. To validate the accuracy of those fitted empirical critical values when SNPs are typed on a real platform, we selected these SNPs based on the SNP information obtained from three different platforms: the Illumina HumanHap550 Genotyping BeadChip, the Human1M BeadChip (<http://www.illumina.com/>) and the Affymetrix Genome-Wide Human SNP Array 6.0 (<http://www.affymetrix.com/>). Using the SNPs that are on these three platforms from different companies, we re-performed the simulation (50 replications) and permutation to obtain the empirical critical values. Table 1 compares the platform specific critical values with those from Figure 3. Under most testing conditions, the fitted critical values based on the nonparametric estimation are close to the average values of the empirical critical values. In other words, our nonparametric estimation as presented in Figure 3 is empirically unbiased and robust to reasonable SNP spacing in common platforms.

Finally, we compared the fitted critical values obtained from nonparametric estimation to the empirical critical values from permutation in the Cancer Genetic Markers of Susceptibility (CGEMS, <http://cgems.cancer.gov/>) breast cancer projects (a balanced dataset with 1,231 case subjects and 1201 control subjects, and 532,075 autosomal SNPs after removal of SNPs with excessive missing rate) (Hunter et al., 2007). We used the CEU population to estimate the fitted critical values because the subjects were of European ancestry (Hunter et al., 2007). Since we had not performed simulations based on 2,434 samples, we obtained the fitted critical values of 30.96 and 5.251 for genotype χ^2 test and Armitage trend test (additive model) by interpolation from the values for 2,000 samples (30.91, 5.248) and 4,000 samples (31.12, 5.264). In comparison, the Bonferroni corrected critical values are 32.40 and 5.341, and the average empirical critical values from permutations are 31.38 (SD = 0.31) and 5.269 (SD = 0.028), respectively. This result further confirmed that our nonparametric estimation of the critical value is close to the average value of the empirical critical values in real datasets.

We have developed a web-based program for estimating the empirical genomewide p values for a GWA study for three common genotyping platforms. The program is available at <http://c2s2.yale.edu/software/gwa-cve>. Features of this program includes the estimation of any critical value at the genomewide significance level of 0.05 for a GWA study based on their population, sample size and the total number of SNPs genotyped.

3.4 Interpretation of the parameters of GEV for genotype χ^2 test

An alternative to estimating the empirical critical values is to estimate the effective number of independent tests in the genome. Once the effective number of independent tests has been estimated, Bonferroni correction could be used to get the critical value at different significance level. Genotype χ^2 test provides an example of how to infer this number from the estimated parameters of the corresponding GEV. There are 3 parameters for a GEV distribution, μ (location parameter), σ (scale parameter) and ξ (shape parameter). In Appendix 1, we show that when the maximum statistic is derived from the genotype χ^2 test, σ is 2 and the number of effective independent tests is $e^{\mu/2}$ (Appendix 1). Table 2 lists the estimated σ , μ , the effective number of independent SNPs, the critical value based on the effective number and its empirical counterpart at different SNP density. Table 2 confirms that the correlation is more severe in high SNP densities. For example, when SNP density is below 300K, the effective number of independent tests increases by about 90% as the SNP density doubles. However, the effective number increased by 56% when SNP density doubles from 1.2M to 2.4M. Table 2 also reveals that the inferred critical value from the estimated effective number of independent tests is slightly larger compared to the empirical critical value. Our estimation of the effective number of independent tests was similar to the number estimated by other methods (Gao et al., 2009; Moskvina and Schmidt, 2008).

4. Discussion

GWA studies have produced landmark successes in identifying genetic variants for complex diseases. We proposed an effective way to deal with one of the most challenging problems of conducting such a GWA study; that is to find the true null distribution of the test statistic and hence to find the proper level of genomewide test significance. We noted the lack of efforts to systematically describe the true null distribution. Instead, the Bonferroni correction and the permutation method are two most popular approaches to derive the significance level. The former is the simplest, but also the most conservative (ignoring the nature of the genome other than the total number of markers tested). The latter is more reflective of the genome, but computationally intensive. In the current report, using simulated data that faithfully follow the LD structure of the input population, we studied the general attributes of the empirical distributions of the test statistics generated under null hypothesis for several association tests

commonly used in GWAs. By taking advantages of the generated null distribution, we are able to quickly and accurately estimate the empirical significance level (see our web-based program) for several commercial genotyping chips, and to implement a new method to estimate any critical value based on experimental design, which is computationally fast (see our web-based program) and reflects the density and basal correlations of the markers.

We noticed that in most situations, the fitted critical value is smaller than the empirical critical value obtained from permutation for Illumina Human550 and Human1M chips and larger than the empirical value for Affymetrix 6.0 array. This observation is consistent with the fact that Illumina chips have better coverage than the Affymetrix chips (Frazer et al., 2007). It also suggests that we could potentially include the coverage in order to improve performance of the fitted models.

In simulating the genomewide genotypes, we purposely avoided using a specific platform. Instead, we used the largest phased data currently available. This precaution guarantees that the fitted critical values could be applied to any genotyping platforms. In our simulation, we realized that the estimated empirical critical values have large variations, which suggested a great degree of uncertainty in a permutation-based approach. Our approach overcomes this problem by fitting nonparametric curves based on many independently simulated data. While our estimated distribution is also subject to uncertainty, the magnitude of the uncertainty is minimal. We demonstrated our fitted model using the Illumina HumanHap550, Human 1M chips and Affymetrix Genome-Wide Human SNP Array 6.0 with independently simulated population samples as well as data from the CGEMS breast cancer study.

In the current project, we performed a more complete study of the null distribution compared to some published research (Dudbridge and Gusnanto, 2008). We considered different aspects of the experimental designs in common GWA studies (population, size, SNP density, different statistics and case/control ratio), studied the null distribution and provided direct inference of the effective number of independent tests from the estimated parameters from the GEV. However, compared to the Dudbridge and Gusnanto study (Dudbridge and Gusnanto, 2008), where the data are obtained from 3000 subjects, our simulation algorithm is based on unrelated HapMap population with a smaller size.

While our proposed method is applicable to address the multiple testing issue in GWA studies, our primary main goal is not to address the multiple comparison problem or conduct a comparative study with existing methods of multiple comparisons. Instead, through extensive analysis of real and simulated data, we observed consistent characteristics of the test statistics under null distribution, which help us better understand the impact of the density of SNPs on the statistical inference.

Acknowledgments

This research is supported in part by grants K02DA017713 and R01DA016750 from the National Institutes on Drug Abuse. We thank “Yale University Biomedical High Performance Computing Center” (NIH grant RR19895) for computational resources. We also thank Dr. Kelly Cho for her comments.

References

- Cauchi S, Meyre D, Durand E, et al. Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value. *PLoS ONE* 2008;3:e2031. [PubMed: 18461161]
- Chen X, Liu CT, Zhang M, Zhang H. A forest-based approach to identifying gene and gene gene interactions. *Proc Natl Acad Sci U S A* 2007;104:19199–19203. [PubMed: 18048322]

- Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics* 1994;138:963–971. [PubMed: 7851788]
- Conneely KN, Boehnke M. So Many Correlated Tests, So Little Time! Rapid Adjustment of P Values for Multiple Correlated Tests. *Am J Hum Genet* 2007;81.
- Donnelly P, Tavaré S. Coalescents and genealogical structure under neutrality. *Annu Rev Genet* 1995;29:401–421. [PubMed: 8825481]
- Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 2008;32:227–234. [PubMed: 18300295]
- Durrant C, Zondervan KT, Cardon LR, Hunt S, Deloukas P, Morris AP. Linkage disequilibrium mapping via cladistic analysis of single-nucleotide polymorphism haplotypes. *Am J Hum Genet* 2004;75:35–43. [PubMed: 15148658]
- Edwards AO, Ritter R 3rd, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. *Science* 2005;308:421–424. [PubMed: 15761121]
- Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851–861. [PubMed: 17943122]
- Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–2229. [PubMed: 12029063]
- Gao X, Becker LC, Becker DM, Starmer JD, Province MA. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet Epidemiol*. 2009
- Haines JL, Hauser MA, Schmidt S, et al. Complement factor H variant increases the risk of age-related macular degeneration. *Science* 2005;308:419–421. [PubMed: 15761120]
- Helgadottir A, Thorleifsson G, Magnusson KP, et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet* 2008;40:217–224. [PubMed: 18176561]
- Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007;316:1491–1493. [PubMed: 17478679]
- Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 2007;39:870–874. [PubMed: 17529973]
- Jenkinson AF. The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quarterly J of the Royal Meteorological Society* 1955;81:158–171.
- Ke X, Taylor MS, Cardon LR. Singleton SNPs in the human genome and implications for genome-wide association studies. *Eur J Hum Genet* 2008;16:506–515. [PubMed: 18197193]
- Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005;308:385–389. [PubMed: 15761122]
- Li C, Li M. GWASimulator: a rapid whole-genome simulation program. *Bioinformatics* 2008;24:140–142. [PubMed: 18006546]
- McPherson R, Pertsemlidis A, Kavaslar N, et al. A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007;316:1488–1491. [PubMed: 17478681]
- Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies. *Genet Epidemiol* 2008;32:567–573. [PubMed: 18425821]
- Nakagawa S. A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behav Ecol* 2004;15:1044–1045.
- Ozaki K, Ohnishi Y, Iida A, et al. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 2002;32:650–654. [PubMed: 12426569]
- Sabatti C, Service S, Freimer N. False discovery rate in linkage and association genome screens for complex disorders. *Genetics* 2003;164:829–833. [PubMed: 12807801]
- Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007;357:443–453. [PubMed: 17634449]
- Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316:1331–1336. [PubMed: 17463246]

- Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007;316:1341–1345. [PubMed: 17463248]
- Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;445:881–885. [PubMed: 17293876]
- Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;100:9440–9445. [PubMed: 12883005]
- Stranger BE, Forrest MS, Clark AG, et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet* 2005;1:e78. [PubMed: 16362079]
- The International HapMap Consortium. The International HapMap Project. *Nature* 2003;426:789–796. [PubMed: 14685227]
- The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–678. [PubMed: 17554300]
- Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004;96:434–442. [PubMed: 15026468]
- Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;316:1336–1341. [PubMed: 17463249]
- Zhang H. Analysis of infant growth curves using multivariate adaptive splines. *Biometrics* 1999;55:452–459. [PubMed: 11318199]

Appendix 1: The parameters in the distribution of the GEV for genotype χ_2 test

The cumulative distribution function (CDF) of a χ_2^2 random variable is: $F_1(x) = 1 - e^{-x/2}$ and the CDF for the maximum statistic of n independent χ_2^2 random variables is $F_n(x) = (1 - e^{-x/2})^n$. For a large x , $e^{-x/2} \approx 1 - e^{-x/2}$, and $F_n(x) \approx e^{-ne^{-x/2}}$. We know that the asymptotic GEV distribution is $F(x) = e^{-(1 + \xi(\frac{x-\mu}{\sigma}))^{-1/\xi}}$, and thus, $-ne^{-x/2} \approx -\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}$, leading to $\ln(n) - \frac{x}{2} \approx \left(-\frac{1}{\xi}\right) \ln\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)$. In the current analysis, the shape parameter ξ is estimated to be around 2E-3. Note that $\ln\left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right) \approx \xi \frac{x-\mu}{\sigma}$, and hence, $\ln(n) - \frac{x}{2} \approx -\frac{x-\mu}{\sigma}$, which suggests $\sigma \approx 2$ and $n \approx e^{\mu/2}$.

Figure 1.

Samples of empirical null distributions. (A) Empirical null distributions of the maximum statistics for a allelic χ^2 test (left), Armitage's trend test for additive model (middle) and genotype χ^2 test (right), estimated from balanced JPT+CHB population datasets with 2,000 samples and 600,000 SNPs. (B) Empirical distributions for CEU (blue), YRI (red) and JPT +CHB (green) populations, estimated from a genotype χ^2 test for balanced datasets with 2,000 samples and 600,000 SNPs. (C) Empirical null distributions for a genotype χ^2 test for balanced JPT+CHB datasets with 2,000 samples and 150,000 (blue), 600,000 (red) and 2,400,000 (green) SNPs. (D) Empirical null distributions for a genotype χ^2 test for balanced JPT+CHB datasets with 1,000 (blue), 2,000 (green) and 4,000 (red) samples with 600,000 SNPs. (E)

Empirical null distributions for a genotype χ^2 test for JPT+CHB datasets with 600,000 SNPs with either 1,000 cases and 1,000 controls (red) or 1000 cases and 2000 controls (blue).

Figure 2.

Q-Q plot and density plot comparing fitted general extreme value distribution with empirical distribution from permutation replicate for a genotype χ^2 test for balanced JPT+CHB datasets with 2,000 samples and 600,000 SNPs. The fitted GEV distribution and empirical distribution are drawn in solid and dashed lines, respectively in the density plot.

Figure 3.

A scatter plots depicting the relationship between the critical values and number of SNPs in GWAs for the CEU (blue circle), YRI (red square) and JPT+CHB (green diamond) populations with 2000 simulated samples in balanced experimental design. Nonparametric regression curves were drawn for the CEU (blue), YRI (red) and JPT+CHB (green) populations. The black dashed line indicated the Bonferroni corrected critical values.

Table 1

Comparison of critical values obtained from the nonparametric estimation, and permutation (referred to as empirical) using the Illumina HumanHap550, Human1M Genotyping BeadChips and Affymetrix Genome-Wide Human SNP Array 6.0 (Affy 6.0)

Population	Chip	Number of Cases: Number of Controls	Actual number of SNPs	Genotype χ^2 test			Armitage trend test (Additive model)		
				Fitted critical value	Average of empirical critical value	SD of empirical critical value	Fitted critical value	Average of empirical critical value	SD of empirical critical value
CEU	Human550	500:1000	518628	31.49	31.79	0.23	5.288	5.290	0.027
		1000:1000		30.88	31.14	0.29	5.244	5.265	0.026
	Human1M	500:1000	797804	32.17	32.34	0.26	5.348	5.343	0.029
		1000:1000		31.47	31.78	0.31	5.306	5.317	0.026
	Affy6.0	500:1000	677416	31.91	31.78	0.38	5.325	5.304	0.029
		1000:1000		31.24	31.11	0.31	5.283	5.267	0.022
YRI	Human550	500:1000	492170	31.71	31.88	0.33	5.314	5.312	0.028
		1000:1000		31.09	31.23	0.24	5.268	5.278	0.023
	Human1M	500:1000	818344	32.60	32.77	0.28	5.397	5.388	0.031
		1000:1000		31.92	32.01	0.24	5.350	5.356	0.023
	Affy6.0	500:1000	740043	32.42	32.44	0.24	5.380	5.363	0.022
		1000:1000		31.76	31.77	0.29	5.334	5.331	0.024
JPT+CHB	Human550	500:1000	483862	31.33	31.37	0.34	5.268	5.267	0.028
		1000:1000		30.71	30.94	0.29	5.231	5.238	0.023
	Human1M	500:1000	755339	31.98	32.04	0.29	5.333	5.332	0.026
		1000:1000		31.38	31.53	0.31	5.292	5.294	0.027
	Affy6.0	500:1000	641166	31.74	31.57	0.26	5.309	5.295	0.024
		1000:1000		31.14	30.96	0.35	5.270	5.261	0.030

Estimated parameters (μ and σ) of the GEV and effective number of independent tests for a genotype χ^2 test for JPT+CHB datasets at different SNP density with 1,000 cases and 1,000 controls

Table 2

SNP number	$\hat{\sigma}$	$\hat{\mu}$	Effective number of independent Tests	Inferred critical value	Mean (SD) of Empirical critical value
2.4×10^6	1.984	27.292	8.440×10^5	33.05	32.95 (0.25)
1.2×10^6	1.987	26.395	5.390×10^5	32.17	32.06 (0.26)
6×10^5	1.978	25.403	3.283×10^5	31.19	31.06 (0.28)
3×10^5	1.964	24.333	1.923×10^5	30.13	29.96 (0.23)
1.5×10^5	1.961	23.172	1.076×10^5	28.97	28.83 (0.29)
7.5×10^4	1.964	21.924	5.764×10^4	27.71	27.56 (0.31)
3.8×10^4	1.963	20.601	2.975×10^4	26.40	26.19 (0.26)
1.9×10^4	1.968	19.235	1.503×10^4	25.06	24.88 (0.25)