Int J Mol Epidemiol Genet 2011;2(2):163-177
www.ijmeg.org /ISSN:1948-1756/IJMEG1006004

Original Article

Effect of genome-wide simultaneous hypotheses tests on the discovery rate

Susana Eyheramendy[1,2], Christian Gieger[1], Maris Laan[3], Thomas Illig[4], Thomas Meitinger[5,6], Erich Wichmann[7,8,9]

[1]Institute of Genetic Epidemiology, Helmholtz Zentrum Munchen, German Research Center for Environmental Health, Neuherberg, Germany; [2]Department of Statistics, Facultad de Matemáticas, P. Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Santiago, Chile; [3]Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Riia 23, 51010 Tartu, Estonia; [4]Unit for Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; [5]Institute of Human Genetics, Helmholtz Zentrum München, Deutsches Forschungszentrum für Umwelt und Gesundheit, Neuherberg, Germany; [6]Institute of Human Genetics, Technische Universität München, Klinikum rechts der Isar, Munich, Germany; [7]Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; [8]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; [9]Klinikum Grosshadern, Munich, Germany.

Received June 30, 2010; accepted April 22, 2011; Epub May 5, 2011; published May 20, 2011

Abstract: An increasing number of genome-wide association studies are being performed in hundreds of thousands of single nucleotide polymorphisms (SNPs). Many of such studies carry on a second stage in which a selected number of SNPs are genotyped in new individuals in order to validate genome-wide findings. Unfortunately, a large proportion of such studies have been unable to validate the genome-wide findings. In this study we aim to better understand how to distinguish the truly associated features from the false positives in genome-wide scans. In order to achieve this goal we use empirical data to look at three aspects that may play a key role in determining which features are called to be associated with the phenotype. First, we examine the usual assumption of a uniform distribution on null p-values and assess whether or not it affects which features are called significant and the number of significant features. Second, we compare the global behavior of the p-value distribution genome-wide with the local behavior at regions such as chromosomes. Third, we look at the effect of minor allele frequency in the p-value distribution. We show empirically that the uniform distribution is not a generally valid assumption and we find that as a consequence strikingly different conclusions can be drawn regarding what we call significant associations and the number of significant findings. We propose that in order to better assign significance to potential associations one needs to estimate the true distribution of null and non-null p-values.

Keywords: Genome-wide association study (GWAS), single nucleotide (SNPs), p-value distribution

Introduction

The recent advances in genome-wide technology and the patterns of genome-wide variation and linkage disequilibrium provided by the International HapMap project [1-3] have facilitated that a large number of genome-wide association studies are now being performed.

Genome-wide studies [4-7] test thousands of features against some null hypothesis, for example, the null hypothesis of no association between a SNP and a phenotype. In such genome-wide studies, only a small number of features are expected to be truly associated with the phenotype. If one considers the usual 0.01 significance threshold at each test, the number of expected false positive will be 0.01m, where m is the number of tests. In the context of the now popular 1000K/500K genotype data [8-11], this number will give approximately 10000/5000 false positives which could be too large or too expensive if one wants to validate the significant results in an independent dataset. One popular way to control for the type I error is to apply a Bonferroni correction to the significance level. The Bonferroni correction assumes that the m tests performed are independent. Therefore, to get a false positive rate not larger than 0.01, one needs to apply the genome-wide threshold of 0.01/m. Much correlation between features in genome-wide studies is expected, which results in the problem that the threshold is usually too conservative, leading sometimes to no significant features found or many features not being considered significant even though they might be.

In the context of current association studies, finding p-values smaller than 10^{-7} has become the landmark for considering the association study successful. Underlying the claim for finding such small p-values is the assumption that the p-values, under the null hypothesis of no association between the phenotype and the SNP, should distribute uniformly. Therefore, the p-value has to be small enough for the researcher to conclude that it is extremely unlikely that that particular p-value was drawn from the uniform distribution of null cases, which would then lead him/her either to the conclusion that the corresponding SNP truly associate with the phenotype or to consider the SNP in a follow-up study. Unfortunately, a large proportion of such follow-up studies have been unable to validate genome-wide findings [12-16].

Theoretically, we expect that null p-values (i.e. p-values obtained from test of association between a phenotype and a SNP that is not associated with the phenotype) distribute uniformly, but in many situations this could not be the case. Deviations from the uniform distribution can occur for many reasons [17]: (1) correlation among the SNPs; (2) correlation among individuals; (3) unaccountable covariates, for instance in a linear model fitting a continuous phenotype as the response and the SNP as one of the covariates; (4) failure to comply with statistical assumptions of the model, e.g. the normally distributed assumption in linear regression models. In the context of genome-wide association studies other sources of deviation from the uniform distribution of null p-values can be stratification [18,19] and the possibility of many true disease loci of small effects [20]. All of these points can occur in genome-wide association data. Hence, it might be more reasonable to estimate the distribution of the p-values before deciding whether a SNP is associated with the phenotype or not. An advantage offered by looking at the distribution of null and non-null p-values is that we can estimate directly the expected proportion of significant features regardless of the value of the p-values.

Permutation tests offer a solution to the problem of dealing with correlated tests, but in the context of genome-wide analysis, these tests are computationally too expensive. Recent studies [21,22] have proposed simulation-based approaches that can approximate the null distribution of the ordered test statistics significantly faster than permutation tests. Also, a method was proposed that is based on numerical integration of the distribution function, which is also faster than permutation testing [23].

The purpose of our work is to better understand how to distinguish between truly associated and non-associated features from the p-values obtained in a genome-wide scan. In order to achieve this goal, we apply alternative methods to the Bonferroni correction to control for Type I Error to several 500K association studies in the KORA500K S3/F3 sample. In particular, we apply the local false discovery rate and the qvalue methods to the p-values obtained from genome-wide association analysis. These methods depend on an estimator for the proportion of null cases and a distribution for the null cases. We assess the influence that common distributional assumptions have in calling features significant or non-significant. In order to determine whether these methods are sensitive to local behavior, we compare the genome-wide behavior of these methods with the local behavior at each chromosome. Lastly, we look at the minor allele frequency (MAF) of the SNPs to discern whether some SNPs based on MAF can be more likely to get to the set of minimal p-values in a genome-wide scan, in which case, it might be reasonable to consider different thresholds for calling significance at the p-values depending on the MAF of the SNPs.

Materials and methods

Methods

To fix the notation, consider m hypotheses:

$$\{H_1^0, ..., H_m^0\}$$

and the corresponding test statistics $X_1, \ldots, X_m$ for which we have obtained m p-values $\{p_1, \ldots, p_m\}$. Let $p_{(1)}, \ldots, p_{(m)}$ be the ordered p-values. Let Φ^{-1} be the inverse of the standard normal

Table 1. Possible outcomes from m hypotheses tests

	Called non-significant	Called significant	Total
Null True	U	V	mo
Alternative	T	S	m1
Total	W	R	m

cummulative distribution and

$$z_i = \Phi^{-1}(p_i).$$

We now introduce two methods to select "significant" p-values: the local false discovery rate [17], and the qvalues [24]. These two methods were motivated by the seminal paper [31] that introduced the false discovery rate (FDR) method. The false discovery rate is the expected proportion of erroneous rejections among all rejections i.e. E(V/R | R > 0) Pr(R > 0) (see **Table 1** for the definition of the variables). Specifically, define

$$k = \max\{i : p_{(i)} \leq \frac{i}{m} q\}$$

and reject

$$H^0_{(1)}, \ldots, H^0_{(k)}.$$

Benjamini and Hochberg [31] showed that when the test statistics are independent the above procedure controls the FDR at level ≤ q. Let F_0 (F_1) be the cummulative density function of the null (non-null) cases.

FDR(z) is estimated as $w_0F_0(z)/F(z)$, where F_0 is assumed standard normal, $F(z) = w_0F_0(z)+w_1F_1(z)$ is the mixture cumulative distribution of null and non-null cases estimated empirically and the proportion of null cases is estimated to be equal to one ($w_0 = 1$). We now define the qvalue and the local false discovery rate.

qvalue: The qvalue method was proposed in [24] as an extension to the FDR. Similarly to the p-values, the qvalues give each feature its own individual measure of significance, but the qvalues take into account the fact that thousands of features are being tested. The qvalue for a particular feature is the expected proportion of false positives incurred when calling that particular feature significant. For instance, if features with qvalues ≤ 0.01 are called significant then among those significant, one should expect 1% of false positives. Mathematically, the qvalue for pi is defined as follows:

$$\hat{q}(p_i) = \min_{t \geq p_i} \hat{FDR}(t),$$

where $\hat{FDR}(t) = \hat{E}(V)/\hat{E}(R)$), V denotes the number of significant features at level t that are truly null and R the number of significant features at level t.

In estimating the qvalues, two assumptions have been made. First, it is assumed that the expected value of the false discovery rate can be estimated as the ratio of the expectations i.e. E(V/R) ≈ E(V) / E(R). Second, it is assumed that the distribution of the p-values for the null cases is uniformly distributed. Efron [17] looked closer at the latter assumption and noted that the distribution of the null p-values could not necessarily be uniform due to potential unaccounted factors (see Introduction above). We describe next the alternative approach described in [17] to address this issue.

lfdr: If the pi are uniformly distributed, then the distribution of the $z_i = \Phi^{-1}(p_i)$ is a standard normal distribution [26]. Efron [17] estimates the distribution of the zi as a mixture distribution with weights given by the proportion of null and alternative features. Specifically, if f denotes the density of the zi, then f can be written as $f = w_0f_0 + w_1f_1$ where f_0 (f_1) represents the density of the null features (alternative features). The mixture density is estimated non-parametrically using for example splines or local polynomial and the null distribution is estimated as a normal distribution where the mean and standard deviation are estimated, for instance, from the central peak of the mixture density. The local false discovery rate of feature i is then defined as lfdr $(z_i) = w_0f_0(z_i) / f(z_i)$, which has an appealing Bayesian interpretation of being the posterior probability of a null case given z_i, Pr(null | z_i).

Dataset and Phenotypes

In order to empirically assess the robustness of the called-significant results we use data from the KORA500K population study which is described below. In Results we make use of two sets of phenotypes: (1) we use data from the addiction to smoking study to assess the influence of the uniform distribution assumption of the null p-values on the called-significant results; (2) we use data from the blood pressure traits study to assess the influence of minor allele frequency (MAF) on the distribution of p-values. These phenotypes are described in what follows.

Dataset

The KORA research platform (Cooperative health research in the Region of Augsburg) has evolved from the WHO MONICA study (Monitoring of Trends and Determinants of Cardiovascular Disease). The KORA genome-wide association study was recruited from the KORA S3 survey, which is a population-based sample from the general population living in the region of Augsburg, Southern Germany. The study participants were of European origin. The participants were examined in 1994/95 by applying standardized examinations that have been described in detail elsewhere [25]. In the KORA S3 study 4,856 subjects, aged 25 to 74 years, have been examined. 3,006 subjects participated in a follow-up examination of S3 in 2004/05 (KORA F3). Informed consent has been given and the study has been approved by the local ethical committee. For the genome-wide KORA S3/F3 500K study we selected 1,644 subjects of KORA S3/F3 and genotyped them using the Affymetrix 500K Array Set. The phenotypes were taken both from KORA S3 and F3.

The genotyping was performed using Affymetrix 500K Array Set consisting of two chips (Sty I and Nsp I). Hybridisation of genomic DNA was done in accordance with the manufacturers standard recommendations. Genotypes were determined using BRLMM clustering algorithm. The overall genotyping efficiency was 98.26%. Before statistical analysis, we performed filtering of conspicuous chips based on quality measures to ensure robustness of association analysis. Only subjects with overall genotyping efficiencies of at least 93% for both chips and at most one discordant call for 50 SNPs being on both chips were included. We performed analysis on this data using the R statistical software (www.r-project.org) and the PLINK software [27]. We filtered SNPs based on minor allele frequency (≥ 0.01) and call rate (≥ 0.1). A total of 395, 912 SNPs were considered in the analysis. In this study we look at some statistical aspects of these association studies.

Phenotype

Addiction to smoking study. Information about smoking behavior was obtained by standardised interview. Current cigarette smokers are defined as those individuals who reported to smoke currently. This group was further divided into regular and irregular smokers. Regular smokers are individuals who smoke usually one or more cigarretes per day and irregular smokers those who smoke on average less than one cigarette per day. Former smokers are defined as individuals who used to smoke but do not smoke currently and never smokers are defined as individuals who have never smoked.

Genome-wide association tests have been performed on two phenotypes defined based on the classification described above, regular smoker and heavy smoker, and also to the Fagerström score. The definition of the three phenotypes that we analyzed are:

- Fagerström: Standard score (that takes values from zero to ten) based on a questionaire to assess whether an individual smokes because of habit or addiction.
- Regular smoker: Is a 0/1 variable, where cases are defined as either regular or former smokers in F3 and control are defined as never smokers neither in S3 nor F3.
- Heavy smoker: Is a 0/1 variable, where cases are defined as regular smokers that smoke more than 20 cigarrets per day in S3 or F3. Controls are defined as never smokers neither in S3 nor in F3.

Blood pressure traits study. In this study we consider three phenotypes, systolic blood pressure, diastolic blood pressure and hypertension. These phenotypes are defined as follows:

- Systolic (SYS) and diastolic (DIA) blood pressures: For each participant blood pressure measurements were recorded twice: in 1994/1995 (KORA S3) and again ten years later 2003/2004 (KORA F3). Each time,

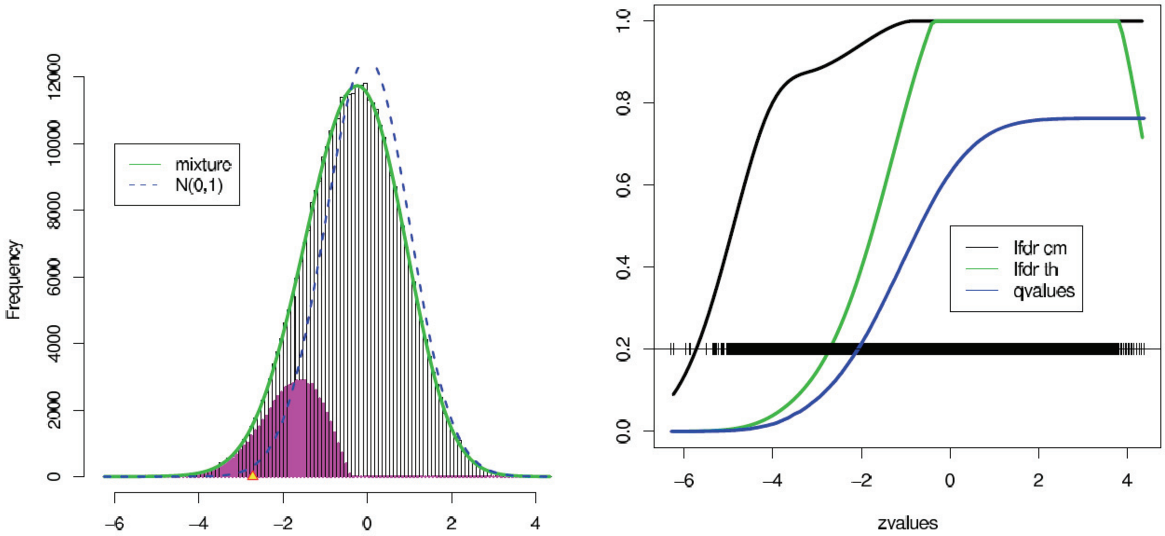

Figure 1. Left: Histogram of z-values for the additive model of the Fagerström phenotype. Blue dashed curve depicts the standard normal distribution and the green curve is the mixture density estimated empirically. The red bars estimate the distribution of the alternative features. Right: Curves for the FDR in red, qvalues in blue and the lfdr with null distribution estimated empirically in black (lfdr cm) or assuming the standard normal in green (lfdr th). The black dashes over the 0.2 horizontal line correspond to the observed z_i values.

systolic and diastolic blood pressures were measured three times at intervals of 3 minutes and the mean of the second and third measurements were taken.

- Hypertension: For the case-control analysis the patients group was formed on the basis of the following criteria: (i) individuals taking antihypertensive medication (during both S3 and F3 survey); (ii) among the patient out of antihypertensive therapy, subjects with SYS ≥ 160 mmHg and/or
- DIA ≥100 mmHg (Grade 2 hypertension) in both S3 and F3 surveys; (iii) untreated individuals with SYS ≥140 mmHg (Grade 1 hypertension) that developed Grade 2 or severe hypertension (SYS ≥ 160 mmHg and/ or DIA ≥ 100 mmHg)
- During the S3 survey, 10 years later. Control subjects were selected to have optimal (≤120/80 mmHg) or normal (120–129/ 80–84 mmHg) blood pressure measured during both surveys, and had never been prescribed antihypertensive medication.

Results

Uniform versus Non-Uniform distribution for the null cases

In this section we use the addiction to smoking data to show that the assumption of the null uniform distribution on the p-values (or equivalently, the assumption of the standard normal distribution on the transformed zi values) affects the number of SNPs called significant in a genome-wide association study.

We fit to each SNP an additive generalized linear Poisson model to the Fagerström score adjusting for age and gender as covariates. The left panel of **Figure 1** shows a histogram of the $z_i = \Phi^{-1} (p_i)$ values. An estimate of the mixture density of null and non-null zi values is depicted in green and the usually assumed standard normal distribution for the null zi is shown as the blue dashed curve. We assume a normal distribution for the null cases zi with unknown mean μ and variance σ^2, which we estimate using the zi lying in the central peak of the mixture distribution. We obtain $\mu = -0.24$ and $\sigma = 1.19$ using the locfdr R package, which considerably deviates from the standard normal distribution. Therefore, we have shown that the null p-values do not necessarily distribute uniformly.

If in this particular example we apply a Bonferroni corrected threshold to the p-values, at 0.01 and 0.05 significance level at each test, then we obtain 19 and 48 significantly associated SNPs respectively. On the other hand, if we estimate the null zi values as suggested by [17], using a normal distribution with mean and stan-

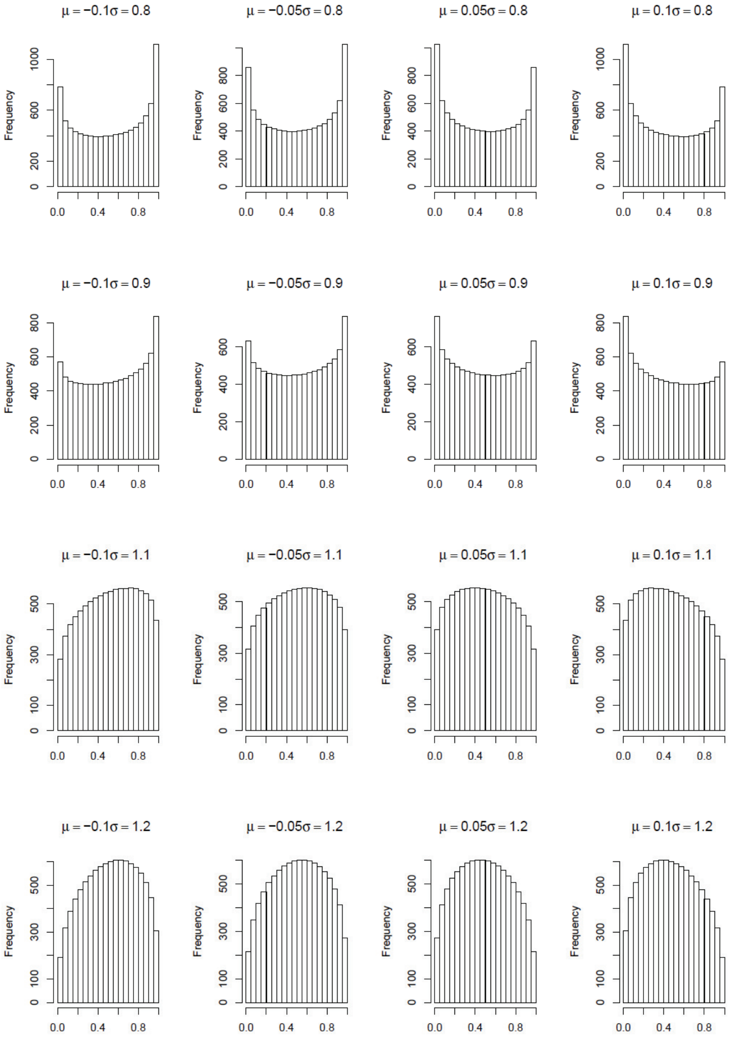

Figure 2. Different shapes for the distribution of null p-values.

dard deviation estimated from the central peak of the mixture distribution, the local false discovery rate at level 0.2 ([17] propose this threshold to be a sensible one) yields 7 significant p-values. The right panel of Figure 1 shows the curve of local false discovery rate in black; the black dashes on the 0.2 horizontal line at the left of this curve correspond to significant results. We show in green the lfdr curve when the null zi are assumed to be distributed as standard normal, giving 8873 significant p-values. Also shown in blue is the curve of the qvalues (giving 7570 qvalues smaller than 0.1). This example shows that the number of significant SNPs can vary significantly if we change the usual assumption on the distribution of the p-values. In particular, note that methods that assume the null-distribution of p-values to be uniform give a much larger number of significant associations.

Finally, logistic regressions were fitted to the regular and heavy smoker phenotypes, adjusting by age and gender as covariates to the additive effect model. As above, we estimate the distribution of null cases with a normal distribution fitted to the central peak of the mixture distribution of null and non-null features. We obtain $\hat{\mu} = -0.035, \ \hat{\sigma} = 1.016$ for the heavy smoker phenotype and $\hat{\mu} = -0.018$ and $\hat{\sigma} = 1.012$ for the regular smoker. Both null distributions do not deviate considerable from the standard normal, and therefore the different methods yield somewhat similar number of significant results. The heavy smoker model yields five SNPs with local false discovery rate below 0.2 when the standard normal is assumed for the null cases and zero when the distribution is estimated by central matching. Four qvalues are below 0.2 and no p-values are below 10^{-7} giving no significant SNPs after Bonferroni correction. At the above mentioned threshold, the regular smoker phenotype gives no significant results using any method.

Theoretical distribution

Consider m hypotheses $\{H_1^0, \ldots, H_m^0\}$ and the corresponding test statistics $X_1, \ldots, X_m$ for which we have obtained m p-values $\{p_1, \ldots, p_m\}$. Let Φ^{-1} be the inverse of the standard normal cumulative distribution and $z_i = \Phi^{-1}(p_i)$. In theory, if the null p-values are independent, their distribution is uniform (i.e. U [0,1]) which is mathematically equivalent to say that the distribution of the $z_i = \Phi^{-1}(p_i)$ follows a standard normal (i.e. N (0, 1) normal distribution with mean zero and variance one).

In this section we revise the consequences that a non-uniform distribution, on the truly not associated SNPs, has on the discovery rate. If we assume that the distribution of the z_i is normal with mean μ and variance σ^2 then the cumulative distribution of the corresponding p-values $p_i = \Phi^{-1}(z_i)$ is given by

$$F(p) = \Pr(p_i \leq p) = \Phi^{-1}(\frac{\Phi^{-1}(p) - \mu}{\sigma}) \quad .$$

This distribution is uniform only in the case that the mean is zero and the variance is one, $\mu = 0$, $\sigma^2 = 1$. **Figure 2** shows different shapes that the distribution of p-values can take. We can see that values of σ above one lead to concave-shape distributions while values of σ below one lead to convex shape distributions. Also, a positive mean leads to a higher concentration of values on the left-hand side of the distribution while a negative mean leads to a higher concentration of values on the right-hand side of the distribution.

For different values of $\sigma \in \{0.8, 0.85, 0.9, 0.95, 1, 1.02, 1.05, 1.1, 1.2\}$ and different values of the mean $\mu \in \{-1, -0.5, -0.2, -0.1, -0.05, -0.02, 0, 0.02, 0.05, 0.1, 0.2, 0.5, 1\}$, we calculate the probability of obtaining a p-value smaller than 10^{-7} ($\Pr(p \leq 10^{-7})$). **Figure 3** shows a levelplot of the log of these probabilities. Lighter regions are regions in which the probability is smaller that 10^{-7} and darker regions are regions for which the probability is higher than 10^{-7}. Parameters leading to darker regions will generate a higher number of false positives than expected under the uniform null distribution while parameters leading to lighter regions will generate a lower number of false positives than expected.

Chromosome-wide versus genome-wide analysis

We have seen in the previous sections that different phenotypes not always have the same distributional properties. If we assume that the LD structure of each chromosome is independent of other chromosome's LD, then considering a single distribution over the whole genome could affect the discovery rate in some regions. In this section we want to assess whether the distribution assumed to hold genomewide is suitable for each chromosome or whether differences can be captured if we look and assess the significance of the associations at each chromosome separately.

Figure 4 shows as black curves the general genome-wide local false discovery rate and qvalue curves for the 1000 best scores of the three smoking related phenotypes. The colored curves show the same quantity computed using selected chromosomes indicated in the legends. This Figure illustrates that the Genome-wide behavior can be very different than the local behavior at each chromosome.

For example, in the panel on the second row and third column the minimum genome-wide value attained by the local false discovery rate (null distribution estimated using central matching) is 0.246 for the heavy smoker phenotype when estimated using z-values. At the threshold level of 0.2 no SNP will appear to significantly associate genome-wide with the heavy smoker phenotype based on the lfdr. Alternatively, if we estimate the lfdr using z-values from chromosome five, we obtained a minimum of 0.011 for the lfdr and 19

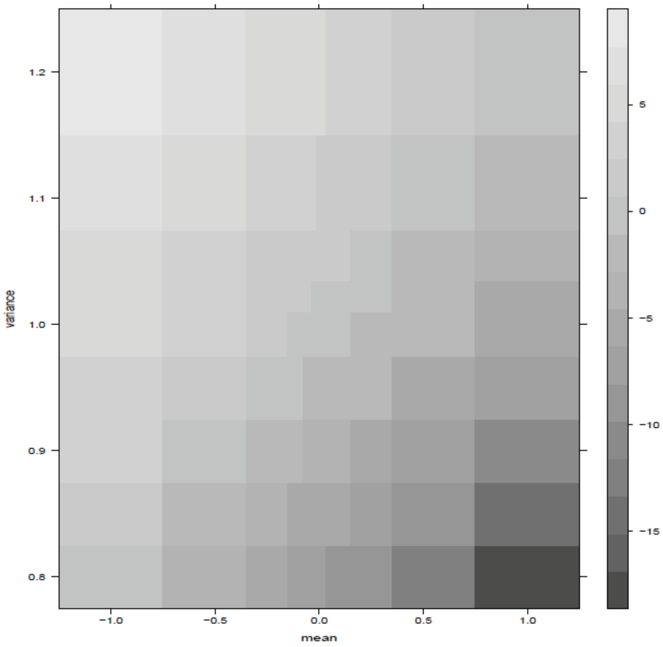

Figure 3. Level plot for the log-probability of obtaining a p-value smaller that 10^{-7} for different values of μ and σ.

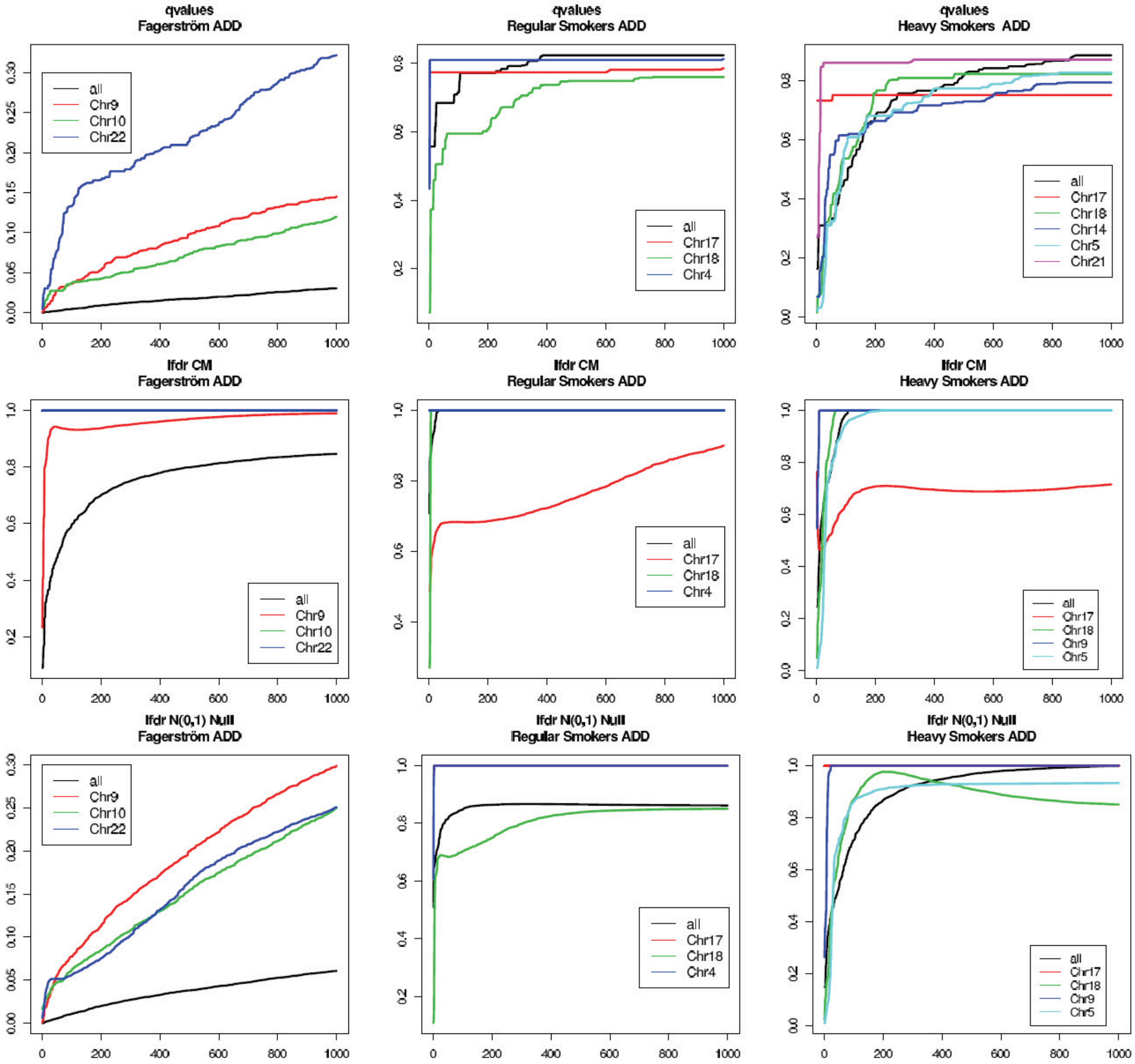

Figure 4. (First row) In black is the curve of the best thousand qvalues genome-wide. In colors are the curves of the best thousand FDR measured chromosome-wide. (Second row). In black is the curve of the best thousand local false discovery rates (null distribution estimated using central matching) genome-wide. In colors are the curves of the best thousand local false discovery rates measured chromosome-wide. (Third row) In black is the curve of the best thousand local false discovery rates (assuming a uniform null distribution) genome-wide. In colors are the curves of the best thousand local false discovery rates measured chromosome-wide. The first, second and third columns corresponds to the Fagerström, regular smoker and heavy smoker phenotypes respectively.

SNPs appear to be significant. In the same way, if we estimate the lfdr using z-values from chromosome 18 we obtain a minimum of 0.048 and 4 significant SNPs. Note that all p-values for the heavy smoker phenotype fitted with the additive model are larger than 10^{-7}, thus genome-wide no p-value is smaller than the Bonferroni corrected threshold.

Examining now the behavior of the qvalues for the heavy and regular smoker phenotyopes, we can see that the relevant curves in **Figure 4** show similar behavior. For the regular smoker phenotype, the smallest qvalues genome-wide is 0.56, yielding no significant SNPs, while the smallest qvalue in chromosome 18 is 0.07, yielding three SNPs with qvalues smaller than

Table 2. Estimates of the proportion of null features using the method described by Storey and Tibshirani24. The p-value correspond to the test that measure whether there are differences in the proportion of null cases computed using genome data or chromosome data.

Fagerström			Regular Smokers			Heavy Smokers		
Chr	Null Prop	P-value	Chr	Null Prop	P-value	Chr	Null Prop	P-value
All	0.763		All	0.972		All	0.976	
9	0.724	3.799e - 41	5	0.926	0	4	0.938	0
10	0.711	1.579e -87	14	0.929	0	17	0.910	0
22	0.713	4.489e -20	21	0.915	0	18	0.936	0

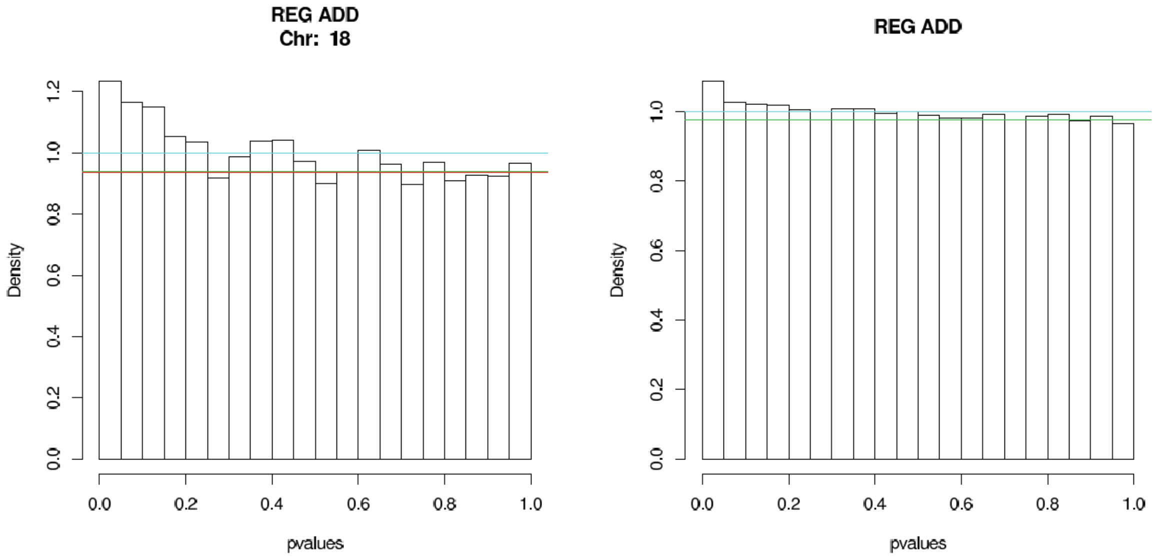

Figure 5. Left panel shows histogram of p-values from chromosome eighteen obtained from the regular smoker phenotype fitted to an additive effects logistic model. The right panel shows histogram of p-values obtained likewise over the whole genome. The light blue line in the graphs represents the uniform distribution that would be observed if all p -values have been obtained from null cases. The green and red lines represent estimated proportions of null features.

0.1. For the heavy smoker phenotype, the smallest qvalue genome-wide is 0.16, while the smallest qvalues for chromosomes 5, 14 and 18 are 0.021, 0.069 and 0.015, giving 23, 9 and 16 qvalues smaller than 0.1 respectively.

As a last example, consider the lfdr with normal assumptions for the heavy and regular smoker phenotypes. For the regular smoker phenotype the lfdr with a standard normal distribution fitted to the null cases gives a minimum value of 0.51 and 0.11 genome-wide and at chromosome 18, yielding zero and three rates smaller than 0.2 respectively. Similarly, the heavy smoker phenotype yields 0.14, 0.01 and 0.02 minimum values for the lfdr with 5, 20 and 13 rates smaller than 0.2 genome-wide, at chromosome 5 and 18 respectively.

A very different behavior can be seen in the corresponding plots for the Fagerström score in Figure 4. The minimum qvalue and local false discovery rates attained genome-wide are significantly smaller than the corresponding values chromosomewide possibly due to the considerably larger proportion of Bonferroni genome-wide significant SNPs in this model.

In the three smoking related phenotypes the particular chromosomes depicted in Figure 4 were selected based on an estimator of the null cases. Chromosomes with the largest difference between the proportion of null cases genome-wide and chromosomewide were selected. As an example, Figure 5 shows in the left panel a histogram of the p-values obtained from chromosome eighteen for the additive model in the regular smoker phenotype, while the right panel shows the corresponding histogram with p-values genome-wide. While the proportion of null cases is estimated as 0.976 genome-wide, at chromosome eighteen is estimated to be only 0.936. Thus, we should expect to see proportionally 4% more significant p-values in chromosome eighteen than genome-wide. As another example, the proportion of null p-values estimated from the Fagerström phenotype model is 0.763 genome-wide while at chromosome 22 and 10 is slightly larger than 0.71. So again, proportionally about 5% more significant p-values are expected in these chromosomes compared to genome-wide.

The same behavior can be seen in the distribution of p-values of all the models that we tested. It is natural that the number of significant p-values will depend on the size of the linkage disequilibrium (LD) region in which the causal SNP lies [28,29], but also on the number of truly associated features. It is important to distinguish between these two possible sources of significant p-values. Table 2 shows estimates of the proportion of null cases genome-wide and in three chromosomes for the smoking related phenotypes. The three chromosomes shown for each phenotype are those with the largest difference of null proportions compared with the corresponding proportion genome-wide. These proportion are estimated using the method described in [24]. The proportions of null features estimated genome-wide and by chromosome are significantly different as shown in Table 2.

p-values distribution in rare and common SNPs

Another aspect that might be influencing the p-value distribution obtained from an association test is the minor allele frequency (MAF) of the SNPs at which the test is performed. We used the blood pressure traits data in order to assess whether or not it is more likely for some SNPs to get into the top smaller p-values obtained from a genome-wide scan based on their MAF. In order to do this, logistic regressions with additive effect adjusted by age and gender were fitted to each SNP for the hypertension phenotype. For systolic and diastolic blood pressures, linear regressions with additive effect were fitted adjusting for age and gender.

Figure 6 shows the distribution of MAF of the SNPs from the KORA population study. In the upper left panel we show the distribution of MAF over the whole range of the p-values. In the other panels we show the MAF distribution for SNPs with p-values smaller than 10^{-4} for the three blood pressure trait phenotypes. While the behavior on the systolic and diastolic blood pressures shows a clear preference for rare SNPs to have smaller p-values, this behavior is not observed for the binary hypertension phenotype.

Naively one might have expected that a rare SNP would have a higher chance of obtaining a smaller p-value. As the KORA blood pressure trait illustrates, this might not always be the case. It will be of interest to explore the level to which a potential association between rare SNPs and high significance using larger amounts of phenotypes and genotype datasets. While such study is beyond the scope of this work, we note that we can already conclude that such an association is not always present.

Discussion

In this study we have shown empirically that the uniform distribution assumption of the null p-values can be an inaccurate estimate of the true distribution of null p-values. We show that as a consequence strikingly different results in the number of called-significant features can be obtained if the distribution is estimated directly from the data.

We also show that the genome-wide behavior does not in general capture local behavior at each chromosome in terms of the proportion of null cases and the distribution of null cases. This is arguably due to differences in the LD structure in different regions [28,29], but could also be because of different proportions of unlinked SNPs associated with the phenotype. While p-values smaller than say 10^{-7} are unlikely to be drawn from the uniform distribution of null cases, an excess of p-values smaller than say 10^{-6} can also be equally unlikely. The large number of SNPs that are being tested simultaneously in genome-wide scans are making it difficult to find new associations due, in part, to the stringent thresholds that need to be applied in order to avoid the multiple testing problem. Therefore, selecting the top smallest p-values from a genome-wide scan that are Bonferroni-significant can leave out many possibly interesting SNPs.

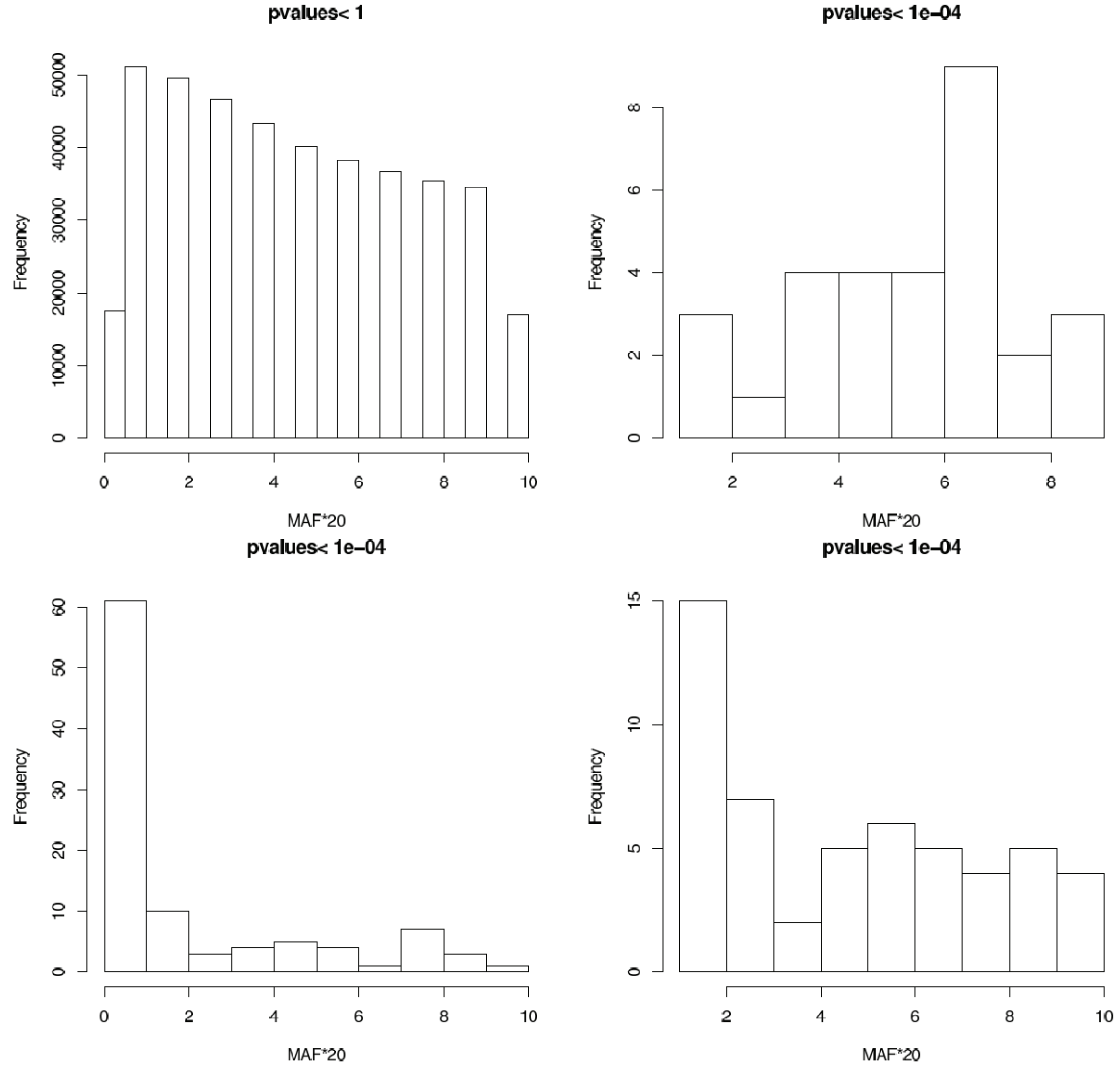

Figure 6. (First row) Histograms of the MAF of all SNPs (left panel). Histograms of MAF for all SNPs with p-values ≤10^{-4} obtained from t-tests, additive effect in logistic regression analysis with HYPertension as phenotype (right panel). (Second row) Histograms of MAF for p-values ≤ 10^{-4} obtained from t-tests, additive effect linear models (for SYS and DIA blood pressures respectively).

Ideally one would know the true distribution of the p-values. Given that this is not generally the case, it is better to adopt an approach that estimates this distribution from the data itself. This is particularly true given that the uniform distribution assumption for the null-cases is not generally valid and is the basis of the most widely used thresholds in genome-wide association studies. We suggest that the failure of that assumption can lead to a non-optimal selection of (associated) SNPs for follow-up studies, and might have played a role in the failure to date to validate significant results.

In order to better assign significance to p-values in genome-wide association studies considering the points raised above, we propose the following procedure.

- Estimate distribution of null and non-null cases using, for instance, the central matching method [17]. This distribution can then be used to: (1) test whether or not the uniform distribution is appropriate as the null distribution, and (2) estimate the proportion of null and non-null features based on the estimated null distribution.

- If the distribution of the null cases does not differ significantly from the uniform distribution (or equivalently, if it does not differ from the standard normal when transformed p-values are used) then select the best features based on genome-wide analysis in the proportion of the expected non-null features as estimated previously (e.g., [24], [29]). Otherwise, if the distribution of the null cases does differ significantly from the uniform distribution then estimate the proportion of null cases using for example the method in [17] or [31].

If the researcher is inclined to select SNPs proportionally by chromosome, then follow the previous steps at each chromosome.

The role of MAF in determining the p-values is a largely unknown factor, but as shown here it is not possible to assume generically that rare SNPs will have smaller p-values. The procedure we suggest has less underlying assumptions. By adjusting better to the observed data it might lead to a higher rate of success in identifying associations in genome-wide studies.

Acknowledgements

S.E. would like to thank the financial aid received from Fondecyt project number 11085012. The KORA Augsburg studies were financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany and supported by grants from the German Federal Ministry of Education and Research (BMBF). Part of this work was financed by the German National Genome Research Network (NGFN). Our research was supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ.

Address correspondence to: Dr. Susana Eyheramendy, Department of Statistics, Facultad de Matemáticas, P. Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Santiago, Chile.E-mail: susana@mat.puc.cl

References

[1] International Hapmap Cpnsortium. A haplotype map of the human genome. Nature 2005;437, 1299-1320.

[2] International Hapmap Cpnsortium , Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, CAI D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy, S. , Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million snps. Nature 2007;449, 851– 861.

[3] Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Ler ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C,

Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Johnson TA, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. Genome-wide detection and characterization of positive selection in human populations. Nature 2007;449, 913–918.

[4] Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, Wallace C, Howson JM, Guja C, Ionescu-Tirgoviste C, Simmonds MJ, Heward JM, Gough SC, Dunger DB, Wicker LS, Clayton DG. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007;39, 857–864.

[5] Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhart AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR. Genome-wide association study identifies new susceptibilityloci for crohn disease and implicates autophagy in disease pathogenesis. Nat Genet 2007;39, 596–604.

[6] Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI. A common variant in the fto gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007;316, 889–894.

[7] Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo KY, Noh DY, Ahn SH, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low YL, Bogdanova N, Schurmann P, Dork T, Tollenaar RA, Jacobi CE, Devilee P, Klijn JG, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, MacPherson G, Reed MW, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko YD, Spurdle AB, Beesley J, Chen X, Mannermaa A, Kosma VM, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BA. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447, 1087–1093.

[8] Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lind-

blad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316, 1331–1336.

[9] Weedon MN, Lettre G, Freathy RM, Lindgren CM, Voight BF, Perry JR, Elliott KS, Hackett R, Guiducci C, Shields B, Zeggini E, Lango H, Lyssenko V, Timpson NJ, Burtt NP, Rayner NW, Saxena R, Ardlie K, Tobias JH, Ness AR, Ring SM, Palmer CN, Morris AD, Peltonen L, Salomaa V, Smith GD, Groop LC, Hattersley AT, McCarthy MI, Hirschhorn JN, Frayling TM. A common variant of hmga2 is associated with adult and childhood height in the general population. Nat Genet 2007;39, 1245–1250.

[10] Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Konig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H. Genomewide association analysis of coronary artery disease. N Engl J Med 2007;357, 443–453.

[11] Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, Jalilzadeh S, Fulda S, Putz B, Eckstein G, Hauk S, Trenkwalder C, Zimprich A, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Peglau I, Eisensehr I, Montplaisir J, Turecki G, Rouleau G, Gieger C, Illig T, Wichmann HE, Holsboer F, Muller-Myhsok B, Meitinger T. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat Genet 2007;39, 1000–1006.

[12] Ioannidis JP, Ntzani E E, Trikalinos TA, & Contopoulos-Ioannidis, D. G. Replication validity of genetic association studies. Nat Genet 29, 306–309 (2001).

[13] Ioannidis, J. P. A. Why most published research findings are false. PLoS Med 2005;2, e124.

[14] Colhoun HM, McKeigue PM & Davey Smith, G. Problems of reporting genetic associations with complex outcomes. Lancet 2003;361, 865–872.

[15] Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG, Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS. Replicating genotype-phenotype associations. Nature 2007;447, 655 –660.

[16] Clarke GM, Carter KW, Palmer LJ, Morris AP, Cardon LR. Fine Mapping versus Replication in Whole-Genome Association Studies. Am J Hum Genet 2007;81, 995–1005.

[17] Efron, B. Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. Journal of the American Statistical Association 2004;99, 96–104.

[18] Sarasua S, Collins J, Williamson D, Satten G, Allen A. Effect of population stratification on the identification of significant single-nucleotide polymorphisms in genome-wide association studies. BMC Proceedings 3, S13 2009; URL http://www.biomedcentral.com/1753-6561/3/S7/S13.

[19] Heiman GA, G. P. Z. J. G. D., Hodge SE. Effect of Population Stratification on Case-Control Association Studies. Human Heredity 2004;58, 30–39.

[20] International Schizophrenia Consortium, W. N. S. J. V. P. O. M. S. P. S. P., Purcell SM. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460, 748–752.

[21] Lin, D. Y. An efficient Monte Carlo approach to assessing statistical significance in genomic studies. Bioinformatics 2005;21, 781–787.

[22] Seaman, S. R. & Muller-Myhsok, B. Rapid simulation of P values for product methods and multiple-testing adjustment in association studies. Am J Hum Genet 2005;76, 399–408.

[23] Conneely, K. & Boehnke, M. So Many Correlated Tests, So Little Time! Rapid Adjustment of P Values for Multiple Correlated Tests. Am J Hum Genet 2007; 81.

[24] Storey, J. & Tibshirani, R. Statistical significance for genomewide studies. Proc Natl Acad Sci 2003;100, 9440–5.

[25] Wichmann, H.E., Gieger, C. and Illig, T. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 2005;67 (Suppl. 1), S26 – S30.

[26] Casella, G. & Berger, R. L. Statistical Inference (Duxbury Press, 1990).

[27] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. Plink: a tool set for whole-genome association and populationbased linkage analyses. Am J Hum Genet 2007;81, 559–575.

[28] Daly M. J., Rioux J. D., Schaffner S. F., Hudson T. J. & Lander E. S. Highresolution haplotype structure in the human genome. Nat Genet 2001;29, 229–232.

[29] Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC,

Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. Nature 2001;411, 199–204. Comparative Study.

[30] Storey, J. D. A direct approach to false discovery rates. Journal of the Royal Statistical Society, Series B: Statistical Methodology 2002;64, 479 –498.

[31] Benjamini, Y. & Hochberg, Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. Journal of Educational and Behavioral Statistics 2000;25, 60–83.