

Published in final edited form as:

Int J Epidemiol. 2009 October ; 38(5): 1364–1373. doi:10.1093/ije/dyp285.

Ranking of genome-wide association scan signals by different measures

Ulf Strömberg^{1,*}, Jonas Björk^{1,2}, Paolo Vineis³, Karin Broberg¹, and Eleftheria Zeggini^{4,5}

¹Department of Occupational and Environmental Medicine, Lund University Hospital, Lund, Sweden.

²Competence Center for Clinical Research, Lund University Hospital, Lund, Sweden.

³Department of Epidemiology and Public Health, Imperial College, London, UK.

⁴Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

⁵Wellcome Trust Sanger Institute, Hinxton, UK.

Abstract

Background—The *P*-value approach has been employed to prioritizing genome-wide association (GWA) scan signals, with a genome-wide significance defined by a prior *P*-value threshold, although this is not ideal. A rationale put forward is that the association signals rather should be expected to give less support for single nucleotide polymorphisms (SNPs) that are rare (with associated low-power tests) than for common SNPs with equivalent *P*-values, unless investigators believe, *a priori*, that rare causative variants contribute to the disease and have more pronounced effects.

Methods—Using data from a GWA scan for type 2 diabetes (1924 cases, 2938 controls, 393 453 SNPs), we compared *P*-values with four alternative signal measures: likelihood ratio (LR), Bayes factor (BF; with a specified prior distribution for true effects), ‘frequentist factor’ (FF; reflecting the ratio between estimated—post-data— ‘power’ and *P*-value) and probability of pronounced effect size (PrPES).

Results—The 19 common SNPs [minor allele frequency (MAF) among the controls >29%] yielding strong *P*-value signals ($P < 5 \times 10^{-7}$) were also top ranked by the other approaches. There was a strong similarity between the *P*-values, LR and BF signals, in terms of ranking SNPs. In contrast, FF and PrPES signals down-weighted rare SNPs (control MAF < 10%) with low *P*-values.

Conclusions—For prioritization of signals that do not achieve compelling levels of evidence for association, the main driving force behind observed differences between the various association signals appears to be SNP MAF. The statistical power afforded by follow-up samples for establishing replication should be taken into account when tailoring the signal selection strategy.

Keywords

Bayes factor; effect size; likelihood ratio; single nucleotide polymorphism; statistical power; statistics

© The Author 2009; all rights reserved.

*Corresponding author. Department of Occupational and Environmental Medicine, Lund University Hospital, SE-221 85 Lund, Sweden. ulf.stromberg@med.lu.se.

Conflict of interest: None declared.

Introduction

Since the advent of genome-wide association (GWA) studies, enormous amounts of data have been generated and, thereby, many potential routes to further identification of genes involved in common and complex human diseases are open. Potential risk genes have primarily been selected for follow-up on the basis of their very high statistical significance (top-ranking variants). A genome-wide significance is typically defined by means of a prior P -value threshold. Consideration of the false-positive report probability¹ leads to a simple mathematical formula for assessing a significance threshold for ‘strong’ signals²:

$$\begin{aligned} &\text{Posterior odds for true association} \\ &= \text{prior odds for true association} \quad (1) \\ &\times \text{power/significance threshold} \end{aligned}$$

This approach simplifies the analysis by assuming a binary choice between ‘true association’ corresponding to a fixed effect size for a single nucleotide polymorphism (SNP) and ‘no association’ (corresponding to the null effect). By interpreting power as the mean power across SNPs—calculated based on the specified, fixed effect size for true association—an *a priori* genome-wide significance threshold is thereby obtained (hence, the variation in power across the SNPs is not taken into account). For example, with the prior odds of a true association at any specified loci of the order of 100 000 : 1 against and 50% power, a P -value threshold of 5×10^{-7} implies that the posterior odds in favour of a strong signal being a true association would be 10 : 1, that is, 10^6 times higher than the prior odds. However, it has been acknowledged that a genome-wide significance is not ideal.² A rationale put forward is that the association signals rather should be expected to give less support for rare SNPs (with associated low-power tests) than for common SNPs with equivalent P -values, unless investigators believe, *a priori*, that rare causative variants contribute to the disease and have more pronounced effect sizes.² In fact, the statistical power at a fixed effect size can be affected markedly by the risk allele frequency. In reporting power for GWA studies, it is a common practice to assume that the causal variant, or an SNP in perfect linkage disequilibrium (LD) with it, is typed. Table 1 demonstrates the strong influence of risk allele frequency on power.

The problem of sorting out false-positive association signals and reducing false-negative ones provides a crucial challenge for GWA studies.³ The process of achieving convincing statistical support for risk alleles from GWA studies includes two important tasks: (i) ranking the association signals, to provide a prioritization list of SNPs for replication studies; and (ii) calibrating inference to make a decision on whether a genetic variant is associated with the disease or not.^{4,5} As we pointed out above, the P -value approach has commonly been used for ranking SNPs in GWA studies. Moreover, consideration of the false-positive report probability [Equation (1)] has provided a calibration method in some studies—the list of ‘strong’ signals has thereby been decided.² Further feasible statistical approaches to accomplishing these tasks exist. The frequentist and Bayesian paradigms offer different statistical frameworks.^{5,6} There are also differences between the statistical tools based on hypothesis testing (addressing the question ‘Is there a significant association?’)^{7,8}; likelihood ratio (LR) (‘What degree of evidence for an association do the observed data provide, considering a specified alternative hypothesis for calibrating inference?’)⁶; Bayes factor (BF) (‘What degree of evidence for an association do the observed data provide, considering a prior distribution for true effect sizes?’)^{4,5} and effect estimation (‘What can we say about the effect size?’).⁹

We here address the question ‘How should investigators take into account the fact that the associations at different SNPs are estimated with a different precision (which relates to

varying power for a true association) when ranking GWA scan signals?' by considering empirical data. Hence, our focus is on ranking of candidate SNPs in terms of evidence, rather than on calibrating inference. We compare association signals provided by P -values with three association signal measures reflecting degree of evidence for an association [LR, BF and 'frequentist factor' (FF)] and one estimation-based measure [probability of pronounced effect size (PrPES)]. Some of the signal measures rely on some form of prior assumption about true effect sizes; we underline that our prior assumptions about potential effect sizes are similar across the candidate SNPs—in particular, we do not assume, *a priori*, that rare causative variants have more pronounced effect sizes than common causative variants.

Materials and Methods

Empirical data

To compare the different types of association signals (described further), we used data from a high-density GWA scan (with Affymetrix GeneChip 500K Mapping Array Set) for type 2 diabetes from the Wellcome Trust Case Control Consortium.^{2,10} This data set incorporates 1924 cases and 2938 controls, with data on 393 453 SNPs passing quality control.¹⁰ Briefly, both samples and SNPs were subjected to quality control. At the SNP level, quality control checks included deviation from Hardy–Weinberg equilibrium ($P < 10^{-4}$ in cases or controls), minor allele frequency (MAF) ($< 1\%$) and call rate (call rate $< 95\%$ for SNPs with MAF $> 5\%$, call rate $< 99\%$ for SNPs with MAF $< 5\%$).

P -values and genome-wide significance

The P -values were obtained from the Cochran–Armitage test for trend.¹¹ The log-additive disease model seems plausible for each potentially influential SNP.¹⁰ The P -value quantifies the discrepancy between observed data and the null hypothesis of no effect (H_0), as the probability of results being as discrepant or more so, given H_0 . Signals with $P < 5 \times 10^{-7}$ have been labelled as 'strong', with consideration to Equation (1) above.² Signals with P -values between 1×10^{-5} and 5×10^{-7} have been labelled as 'moderate' evidence of association.² In our present investigation, we go further down the significance distribution and report the association signals for the SNPs with $P < 0.001$.

LR signals

For each SNP, we used a logistic regression model for the likelihood $\Pr(\text{data} | \gamma, \mu)$ where the log-odds for the disease is equal to $\mu + \gamma Z_i$; Z_i denotes the genotype (coded as 0, 1 or 2) for individual i . The effect parameter γ represents the per-allele change in the log-odds of the disease, predicted towards the minor allele. The parameter μ represents the baseline log-odds of the disease (taking into account that the number of cases has been elevated artificially in the case–control design). The LR is obtained from the observed association data and takes the form

$$\frac{\Pr(\text{data} | \gamma = \widehat{\gamma}, \mu = \widehat{\mu})}{\Pr(\text{data} | \gamma = 0, \mu = \widehat{\mu})},$$

where $\widehat{\gamma}$ is the maximum likelihood estimate of γ and $\widehat{\mu}$ is the estimated baseline log-odds. An LR is calculated by the ratio of densities at $\widehat{\gamma}$ and 0, respectively, from a normal distribution with mean $\widehat{\gamma}$ and standard deviation (SD) equal to the standard error (SE) of $\widehat{\gamma}$.¹²

An LR signal for a candidate SNP is given by $\log_{10}(\text{LR})$. Clearly, we obtain more evidence against the null hypothesis in favour of the conventional (maximum likelihood) effect size estimate by increasing the LR signal.

BF signals

The BF has been proposed as an alternative to P -values for prioritization of association signals in GWA studies.^{2, 4, 5} The BF is obtained from the observed association data and, here, takes the form $\Pr(\text{data}|M_1)/\Pr(\text{data}|M_0)$ [rather than the conventional, inverted form $\Pr(\text{data}|M_0)/\Pr(\text{data}|M_1)$]^{4, 5}; $\Pr(\text{data}|M_k)=\int\Pr(\text{data}|\theta_k, M_k)d\theta$, where $\theta=(\gamma, \mu)$ denotes the parameters for prior model M_k ($k=0, 1$). Bearing in mind that we restrict attention to the log-additive effect model, M_1 denotes a statistical model reflecting the assigned prior distribution for the effect sizes of every copy of a given allele on the log-odds of the disease, conditional on the existence of an association. We use $\mathcal{N}(0, 0.2)$ prior distribution on the effect parameter γ , which has been suggested.² We reinforce that we focus on a specific BF, with the same effect size model M_1 for all SNPs. M_0 denotes the statistical model under the null effect ($\gamma=0$). Moreover, calculation of the BFs requires a prior distribution on the baseline log-odds parameter μ ; we use $\mathcal{N}(0, 1)$ distribution for μ , which have been suggested (the resulting BFs are relatively insensitive to this choice of prior).²

A BF signal is given by $\log_{10}(\text{BF})$. A larger BF signal indicates more evidence against the null effect model M_0 in favour of the alternative model M_1 . The BF signals were obtained from the software package SNPTTEST (<http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html>).

FF signals

Wacholder *et al.*,¹ who proposed consideration of the false-positive report probability, primarily addressed the *a priori* choice of an appropriate significance threshold for classifying noteworthy association signals, Equation (1). They also described how to calculate the false-positive report probability based on observed data. Such post-data calculation is performed by substituting the observed P -value in place of the predetermined significance threshold and recalculating the power for the predetermined alternative effect size but with the precision (SE) of the obtained effect (γ) estimate. Hence, this post-data ‘power’ is conceptually different from the conventional power calculated for predetermined significance level (not the observed P -value), sample size and SNP MAF (as in Table 1). Based on the genotype counts for the cases and controls, respectively, for each candidate SNP, the ratio

$$\frac{\text{Post-data 'power'}}{\text{p-value}},$$

referred to as the FF, can be used for ranking SNPs. We reinforce that the alternative effect for the post-data ‘power’ calculations should be fixed *a priori*.¹ For demonstration, we estimated the post-data ‘power’ for two alternative effect sizes: (i) e^{γ} [i.e. per-allele OR predicted towards the minor allele]=1.15 and (ii) $e^{\gamma}=1.30$.

The FF can be viewed as a BF variant and has also been referred to as the false-discovery rate BF.¹³ The FF is similar, but not identical, to the BF, as thoroughly discussed by Wakefield.⁴

An FF signal is given by $\log_{10}(\text{FF})$.

PrPES signals

PrPES signals are obtained from a semi-Bayes estimation-based procedure.⁹ First, the log-additive effect size of each candidate SNP, together with its 95% confidence interval (CI), is estimated conventionally (i.e. without specifying any prior distribution on γ). Secondly, a semi-Bayes method is used; adjusting the conventional effect size estimates.¹⁴ Such adjustments pull outlying effect size estimates towards the null effect and lead to narrower 95% CIs than with the conventional estimation method.^{9, 14, 15} Conventional effect size estimates that are biased away from the null can be expected in GWA studies, in particular for rare variants with low P -values.¹⁶ Thirdly, the PrPES signal is calculated for each candidate SNP, based on a pronounced effect size (PES) magnitude, which should be specified *a priori*. An appropriate PES magnitude should be related to the prior assumption about the effect size variability across all SNPs, which is used for the semi-Bayes method (as described below). We considered two reasonable PES magnitudes: (i) $e^{\gamma} > 1.15$ (i.e. e^{γ} , predicted towards the minor allele, > 1.15 or $< 1/1.15 = 0.870$) and (ii) $e^{\gamma} > 1.10$. (The reason for the discrepancy between our choice of the latter PES magnitude and our alternative (ii) $e^{\gamma} = 1.30$ used in the FF approach is further addressed in Discussion.)

The PrPES approach relies on three prior assumptions/specifications. First, the true effects of the SNPs are assumed to be exchangeable *a priori*. The exchangeability assumption implies that we offer about the same prior guess with about the same effective sample size for all SNP-specific effect size parameters, $\gamma_1, \gamma_2, \dots, \gamma_N$ ($N=393\,453$ in our application).¹⁷ This assumption essentially means that we offer an uncertain and qualitative prior guess about the similarity of the effects of the SNPs, without saying that effect size parameters are really equal; without seeing the data, we cannot grade the SNP-specific effect sizes.¹⁷ Secondly, the true values on the effect size parameters $\gamma_1, \gamma_2, \dots, \gamma_N$ are approximately normally distributed centred at the overall average effect. Thirdly, a prior variance of the true values on $\gamma_1, \gamma_2, \dots, \gamma_N$ is specified; this variance is denoted VT . VT should not exceed the ‘observed’ sample variance of the effect parameter estimates, VO (Appendix 1). We point out that VT should be much lower than the variance specified for the true effect model, M_1 , in the BF approach [i.e. variance 0.04 (0.2^2)], as M_1 is specified conditional on the existence of an association. Our primary choice on $VT=0.00155$, which implies that we, *a priori*, expect a fraction of 0.36×10^{-5} of the true individual ORs (e^{γ}) to fall > 1.20 or $< 1/1.20 = 0.833$, provided a null overall average effect. In some respect, that choice of VT is in line with the indicated prior viewpoints in the present GWA study: a probability of the genome-wide existence of an association in the order of $100\,000 : 1$ against and, moreover, $N(0,0.2)$ prior distribution on γ —yielding $e^{\gamma} > 1.20$ or < 0.833 with probability 0.36 —conditional on the existence of an association.² We also performed sensitivity analyses by varying VT .

Technical details for the PrPES approach are given in Appendix 1.

Results

***P*-values**

Out of the 393 453 candidate SNPs, 651 have $P < 0.001$. We observe 19 strong signals, that is, with $P < 5 \times 10^{-7}$. These 19 SNPs are common in the study population, with MAFs among the controls $> 29\%$.

LR signals

The LR signals yield virtually the same picture—reflecting the degree of evidence against the null hypothesis across the 651 SNPs examined—as the P -values (Figure 1a).

BF signals

There is also very close agreement between the P -values and BF signals (Figure 1b). Thus, no substantial difference in the primary prioritization list (based on P -values) is provided by the BF approach.

FF signals

For SNPs with equivalently low P -values, the FF signals down-weight rare SNPs (Figure 1c and d), due to the fact that the effect estimates for rare SNPs have relatively weak precision and, therefore, lower post-data ‘power’. Naturally, the usage of the larger alternative effect ($e^{\lambda}=1.30$) implies higher post-data ‘power’ and therefore provides stronger FF signals (Figure 1c vs d).

PrPES signals

The conventional e^{λ} estimates across the 393 453 candidate SNPs range between 0.522 and 1.879, with an observed sample variance (VO ; Appendix 1) equal to 0.003. With the prior variance $VT=0.00155$, the semi-Bayes adjusted e^{λ} estimates range between 0.909 and 1.150. Outlying estimates for rare variants were pulled notably more towards the null effect than outlying estimates for more common variants; for candidate SNPs with control MAF up to 10%, the semi-Bayes adjusted e^{λ} estimates range between 0.940 and 1.081.

The PrPES signals also imply the same list of top ranking variants (Figure 1e and f). The PrPES signals for the higher PES magnitude ($e^{\lambda}>1.15$) are generally weak and yield only marginal down-weighting of rare variants (Figure 1e). For the lower PES magnitude ($e^{\lambda}>1.10$), the PrPES signals are naturally stronger and yield marked down-weighting of rare variants (Figure 1f). Varying PES magnitude had some influence on the ranking—Spearman’s rank correlation coefficient $r_S=0.806$ between the PrPES signals ($n=651$) for the two alternative PES magnitudes at $VT=0.00155$.

The sensitivity analyses confirmed that PrPES signals for the 651 SNPs (with $P<0.001$) were strengthened by increasing VT . We found that the order of the PrPES signals was affected marginally by varying VT (r_S 0.966 when varying our primary choice $VT=0.00155$ by $\pm 20\%$, with $e^{\lambda}>1.15$ as the PES magnitude; and r_S 0.979, with $e^{\lambda}>1.10$ as the PES magnitude). Even an extreme prior assumption about effect variability did not change the ranking substantially (e.g. $r_S=0.919$, $VT=VO=0.003$, vs $VT=0.00155$, with $e^{\lambda}>1.10$ as the PES magnitude).

Robustly replicating loci and examples of SNPs with notably different results

There are now a few robustly replicating type 2 diabetes loci ($n=19$: rs864745; rs12779790; rs7961581; rs7578597; rs4607103; rs10923931; rs10946398; rs5015480; rs10811661; rs757210; rs4402960; rs13266634; rs7901695; rs5219; rs1801282; rs10830963; rs10010131; rs2237892 and rs8050136).¹⁸ We have checked if these SNPs were included in the association signals considered in our present investigation (i.e. among the 651 SNPs with $P<0.001$); seven SNPs in the list above were included (Table 2). Those seven robustly replicating SNPs are common in the study population, with MAFs among the controls $>16\%$. Two robustly replicating SNPs were observed among the 19 strong signals ($P<5\times 10^{-7}$) in this original GWA scan (rs7901695 and rs8050136; Table 2). For the other included SNPs that did not achieve strong evidence for association, but which have gained increased evidence by replication studies, the various signal measures did not give dramatically different support (Table 2; we address differences in results below). The remaining 12 SNPs listed above were not in LD with any of the 651 SNPs with $P<0.001$.

Table 2 also gives examples of rare SNPs (rs657317 and rs12086219) that were notably down-ranked by using the FF (with the alternative $e^{\gamma}=1.15$) and PrPES (with the PrPES magnitude $e^{\gamma}>1.15$) approaches, as compared with the P -value, LR and BF approaches. Hence, the FF and PrPES approaches could sort out rare SNPs from a prioritization list provided by ranking P -values, LR or BF signals. Moreover, there are examples of SNPs for which the FF and PrPES signals, respectively, yield notably different ranks (Table 2). For the specified alternative hypothesis ($e^{\gamma}=1.15$), the FF approach yields less support (in terms of ranking) than the PrPES approach for a SNP with a conventional effect size estimate evidently >1.15 (or $<1/1.15=0.87$); yet, the semi-Bayes adjusted effect estimate provides relatively strong support for a pronounced effect (rs10806665; Table 2). In contrast, the FF approach yields stronger support than the PrPES approach for a SNP with a conventional effect size estimate ~ 1.15 (or $1/1.15=0.87$) (e.g. rs864745; Table 2). One should bear in mind that many association signals by each different measure are dense (Figure 1); in particular, signals by each measure are very similar whenever they relate to very similar likelihood-based estimation results of the effect parameter γ (ignoring direction of effect).

Discussion

Our investigation, based on a single GWA data set, indicates that the selection of the top ranking variants—the common SNPs yielding strong signals—is robust against the choice of signal measure. For prioritization of signals that do not achieve compelling levels of evidence for association, the main driving force behind observed differences between the various association signals appears to be SNP MAF. Rare SNPs with low P -values were down-weighted by using the FF and PrPES approaches; so far, these approaches have not been favoured in practice. The requirement for specifying an alternative in some form, which is not needed for the calculation of a P -value, may have been a practical obstacle. Nevertheless, the statistical power afforded by follow-up samples for establishing replication should be taken into account when tailoring the signal selection strategy.

Performance of the LR approach

The LR approach relies on sound evidential framework theory, dictated by the Law of Likelihood.⁶ For calibrating inference, the evidential likelihood approach means that the null hypothesis and a simple alternative hypothesis (with a fixed effect size) are considered in the design stage and then error probabilities are calculated in order to find a cut-off for the likelihood ratio.⁶ Notwithstanding that aspect, for ranking the association signals after data collection (as we focus on in this article), the LR and P -value approaches can be expected to give similar results. The P -values from the Cochran–Armitage test for trend and the likelihood ratio test (with the null hypothesis $\gamma=0$) are asymptotically equivalent.^{11,12}

Performance of the FF approach

The FF approach simplifies the analysis by restricting true associations to a fixed effect size. The FF signals have conceptual drawbacks and theoretical deficiencies because the calculation ignores information by conditioning on the observed P -value.^{4,19} We point out that the post-data ‘power’ (for a specified alternative)—the numerator of the FF—is lower for an SNP with larger effect size estimate than for an SNP with equal precision but a lower effect size estimate [because the P -value (significance threshold) is lower for the former SNP]; that behaviour of a ‘power’ seems irrational. Nevertheless, we think that FF signals merit consideration in GWA scans due to their straightforward performance: down-weighting rare SNPs at a fixed P -value.

Performance of the BF approach

The BF approach relies on sound statistical theory taking into account some reasonable prior belief about the effect size distribution. We think that this Bayesian framework is attractive for prioritization of association signals. However, we found that rare SNPs with low P -values were not down-weighted by the suggested BF approach (under the given prior model M_1 for the effect parameter, which is independent of SNP MAF). Other GWA scan data support our finding (supplementary figure 22 presented in ref.²). Recently published theoretical results reveal that BF signals at a fixed P -value could, in a well-powered GWA study, give ‘increasing’ support for an association by ‘decreasing’ SNP MAF (implying decreasing precision of the effect estimate).²⁰ On the one hand, the BF approach could yield down-weighting of common SNPs at a fixed P -value because the evidence for alternative effect model M_1 might not be strong—although the data are unlikely under the null effect model M_0 , they are unlikely under M_1 also.²⁰ On the other, the P -value approach under small data departures from the null hypothesis yields small P -values, provided a high precision of the effect estimate. In a well-powered GWA study, the data might yield pronounced (conventional) effect size estimates with fairly high precision, implying only minor bias away from the null,¹⁶ for some influential, rare SNPs. It appears that the BF approach starts to up-weight the support for rare SNPs, relative to common SNPs with equivalent P -value, at a certain sample size.²⁰ However, the sample size required for up-weighting of rare SNPs is likely to be sensitive to the prior specification of the effect size model M_1 (here, assumed to be independent of MAF).

The suggested BF approach could theoretically yield marked down-weighting of rare SNPs at a fixed P -value (if the corresponding effect size estimates are far away from the null). Such results seem to be exceptional in GWA studies of similar size as we considered [we did not observe marked down-weighting of any rare SNP with $P < 0.001$ by the BF approach, although it has been observed for a few rare SNPs in a GWA scan with the same controls but another case series (supplementary figure 22 presented in ref.²)].

Considerations concerning the PrPES approach

A PrPES signal is directly linked to the effect size estimates obtained after the semi-Bayes adjustments of the conventional effect size estimates. Importantly, the pronounced (conventional) effect size estimates are first adjusted downward (towards the null) by the semi-Bayes method. Downward adjustment of extreme, conventional effect estimates is justified, because the estimation results for claimed positive associations are affected by a selection/ascertainment bias referred to as ‘winner’s curse’ (upward bias).^{16,21–23} Shrinkage of the conventional effect estimates by ‘regression towards the mean’ has therefore been advocated for designing replication studies.²² Hence, a positive feature of the PrPES approach is that investigators obtain useful information for specifying reasonable effect sizes when dimensioning replication studies. (The Bayesian approach used for calculating BF signals can also yield simple effect size estimate along with credibility intervals.)

Nevertheless, the calculation of PrPES signals raises some issues of concern. The semi-Bayes method relies on the prior belief about the effect sizes, given by VT under the exchangeability assumption (subsequently, we address modifications of the exchangeability assumptions). In our present investigation, with $VT=0.00155$ as the primary choice, the semi-Bayes adjustments yielded a strongly pronounced stabilization of outlying effect estimates (towards the null effect) for rare SNPs with low P -values. We found that the ranking of the PrPES signals was fairly insensitive to the choice of VT . An additional concern is the prior choice of the PES magnitude—the alternative effect size in the FF approach seems to be an analogous choice. One should bear in mind, however, that an appropriate PES magnitude should be related to the prior assumption about the effect size

variability, VT . For example, an effect size of $e^{\gamma}=1.30$, which has been considered in statistical power calculations for GWA scans,² provides an overly optimistic PES magnitude for our primary $VT=0.00155$, because the expected effect sizes are of lower magnitudes. We advocate the use of a low PES magnitude, which should imply marked variation of PrPES signals.

Rare variants and prior belief modifications

Rare variants are likely to play an important role in complex disease susceptibility. However, for the purposes of prioritizing genome-wide scan association results for follow-up genotyping, we strongly believe that it may be preferable to at least initially pursue common variant associations. The reason behind this rationale is that lower minor allele frequencies do affect power (given modest effect sizes). Replication of such signals would necessitate difficult-to-achieve large numbers of follow-up samples, which most researchers do not have readily available for first-pass replication efforts.

In order to obtain stronger support for rare SNPs, investigators need a better-powered GWA study, different analytical approaches,²⁴ or, alternatively, to put forward functional arguments for giving stronger *a priori* support to rare SNPs. It is possible for investigators to incorporate prior belief modifications, that is, different *a priori* weights to different classes of SNPs, in the various prioritization approaches. By considering Equation (1), different P -value thresholds—and different implications of FF signals—follow from different *a priori* weights across the candidate SNPs. In the PrPES approach, it is possible to modify the exchangeability assumption by assigning different VT s to different classes of SNPs [e.g. considering if they are non-synonymous SNPs, genic SNPs (in particular in genes with a potential function for the disease), SNPs in highly conserved regions or SNPs in LD with many (or few) other SNPs], although such modifications have not been elaborated in practice.⁹

In the BF approach, various choices of the prior model for true associations (M_1) have been suggested.²⁰ An alternative model M_1 proposed implies larger effect size at rarer MAFs.²⁰ However, such an ‘effect-MAF dependence’ prior is questionable as a general rule, without considering functional arguments. A somewhat different M_1 is the ‘implicit P -value’ prior, also for stronger effect sizes to be anticipated at lower MAFs—this prior is also dependent of the sample size, which is undesirable.²⁰ The ‘implicit P -value’ prior is theoretically interesting because it yields identical rankings between (approximate) BFs and P -values.²⁰

Analogously to the prior modifications suggested above for the PrPES approach, it is, in the BF approach, possible to use a different prior variance (specified for the effect size model M_1) for different classes of SNPs.² Such modification of the specific prior model M_1 in the BF approach that we have considered here (i.e. the same M_1 for all SNPs) might be more attractive than modifying M_1 by considering SNP MAF solely.

Two remarks

We focused on the log-additive effect model. The approaches can be generalized to test/fit disease models covering deviations from log-additivity.² We point out that the PrPES approach can be modified by considering the effect size corresponding to a dominant or recessive disease model.

Finally, we stress that our results do not ‘validate’ the different methods, which would require in-depth examination on the performances of the methods—through simulation studies under various scenarios with the real pronounced effects in SNPs with high or low allele frequencies. It is also of interest to study consistency of the different signal measures based on data from replication studies.

KEY MESSAGES

- There are alternatives to the commonly used P -value signal measure that should be considered in practice.
- The main driving force behind differences in performance between the various signal measures appears to be SNP MAF; the top ranking variants (common SNPs) are robust against the choice of signal measure.
- Investigators should take into account the statistical power afforded by follow-up samples for establishing replication when tailoring the signal selection strategy.

Acknowledgments

This work was partly conducted within the EU Network of Excellence ECNIS (<http://www.ecnis.org>). E.Z. is a Wellcome Trust Research Career Development Fellow. The authors thank Dr Jon Wakefield and two reviewers for valuable comments and suggestions.

Funding Swedish Council for Working Life and Social Research [2007-0153 to U.S., J.B. and K.B.]; the Swedish Research Council [2007-22238-47298-36 to J.B., U.S. and K.B.]; the Swedish Cancer Fund [080401 to U.S., J.B. and K.B.] and the Wellcome trust [WT088885/Z/09/Z].

Appendix 1

A Bayesian shrinkage estimator in this context can be defined as²⁵

$$\widehat{\gamma}_{j,adj} = (1 - B_j) \cdot \widehat{\gamma}_j + B_j \cdot \widehat{\gamma}.$$

where B_j is the shrinking coefficient ($0 < B_j < 1$), $\widehat{\gamma}_{j,adj}$ is the adjusted (shrunk) estimate for SNP j , $\widehat{\gamma}_j = \log \widehat{OR}_j$ is the conventional effect estimate for SNP j and $\widehat{\gamma}$ the estimated mean effect of all SNPs. In a semi-Bayes procedure, the shrinkage coefficient is¹⁴

$$B_j = \frac{\widehat{V}}{VT + \widehat{V}_j}$$

where \widehat{V}_j is the estimated variance (squared standard error) of $\widehat{\gamma}_j$ and VT is the prior variance of the true effects γ across all SNPs, assumed to be normally distributed. The expression for the adjusted estimate for SNP j can therefore be written as

$$\widehat{\gamma}_{j,adj} = \frac{VT}{VT + \widehat{V}_j} \cdot \widehat{\gamma}_j + \frac{\widehat{V}_j}{VT + \widehat{V}_j} \cdot \widehat{\gamma}.$$

In other application areas, $\widehat{\gamma}$ is estimated as a weighted average (with w_j as weights) of the conventional effect estimates $\widehat{\gamma}_j$. In GWA scans, however, it is reasonable to set $\widehat{\gamma} = 0$. Thus, the previous expression simplifies to

$$\widehat{\gamma}_{j;adj} = \frac{VT}{VT + \widehat{V}_j} \cdot \widehat{\gamma}_j$$

The observed variance of the distribution of the conventional effect estimates across all SNPs is estimated as

$$VO = \frac{\sum_j w_j \cdot (\widehat{\gamma}_j - \widehat{\gamma})^2}{\sum_j w_j}$$

where

$$w_j = \frac{1}{VT + \widehat{V}_j}$$

since $\widehat{\gamma} = 0$, this expression simplifies to

$$VO = \frac{\sum_j w_j \cdot \widehat{\gamma}_j^2}{\sum_j w_j}$$

The prior variance VT should be chosen such that $VT < VO$. The huge number of SNPs typically involved in GWA scan implies that the variance of the semi-Bayes adjusted effect estimate $\widehat{\gamma}_{j;adj}$ can be approximated by the asymptotic Bayes posterior variance²⁵

$$\widehat{V}_{j;adj} = (1 - B_j) \cdot \widehat{V}_j = \frac{VT}{VT + \widehat{V}_j} \cdot \widehat{V}_j$$

With a prior and conventional effect estimates that are approximately normally distributed, the posterior probability of pronounced effect can be calculated as

$$1 - \Phi \left(\frac{[\log(OR_{high}) - \widehat{\gamma}_{j;adj}]}{\sqrt{\widehat{V}_{j;adj}}} \right) + \Phi \left(\frac{[\log(OR_{low}) - \widehat{\gamma}_{j;adj}]}{\sqrt{\widehat{V}_{j;adj}}} \right)$$

where Φ is the cumulative density function of the standard normal distribution, and OR_{high} and OR_{low} are the given effect sizes (e.g. $OR_{high} = 1:10$ and $OR_{low} = 0:909$).

References

1. Wacholder S, Chanock S, Garcia-Closas M, El-gormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst.* 2004; 96:434–41. [PubMed: 15026468]
2. 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; 447:661–78. [PubMed: 17554300]

3. Vineis P, Brennan P, Canzian F, et al. Expectations and challenges stemming from genome-wide association studies. *Mutagenesis*. 2008; 23:439–44. [PubMed: 18765424]
4. Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet*. 2007; 81:208–27. [PubMed: 17668372]
5. Wakefield J. Reporting and interpretation in genome-wide association studies. *Int J Epidemiol*. 2008; 37:641–53. [PubMed: 18270206]
6. Strug LJ, Hodge SE. An alternative foundation for the planning and evaluation of linkage studies. *Hum Hered*. 2006; 61:166–88. [PubMed: 16865000]
7. Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B*. 1995; 57:289–300.
8. Moerkerke B, Goetghebeur E. Selecting “significant” differentially expressed genes from the combined perspective of the null and the alternative. *J Comput Biol*. 2006; 13:1513–31. [PubMed: 17147475]
9. Strömberg U, Björk J, Broberg K, Mertens F, Vineis P. Selection of influential genetic markers among a large number of candidates based on effect estimation rather than hypothesis testing: an approach for genome-wide association studies. *Epidemiology*. 2008; 19:302–8. [PubMed: 18300718]
10. 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; 316:1336–41. [PubMed: 17463249]
11. Clayton, D. Population association. In: Balding, DJ.; Bishop, M.; Cannings, C., editors. *Handbook of Statistical Genetics*. Wiley; New York: 2003. p. 939–60.
12. Clayton, D.; Hills, M. *Statistical Models in Epidemiology*. Oxford Science Publications; Oxford: 1993.
13. Katki HA. Evidence-based evaluation of P -values and Bayes factors (Invited commentary). *Am J Epidemiol*. 2008; 168:384–88.
14. Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. *Arch Environ Health*. 1994; 49:9–16. [PubMed: 8117153]
15. Steenland K, Bray I, Greenland S, Boffetta P. Empirical Bayes adjustments for multiple results in hypothesis-generating or surveillance studies. *Cancer Epidemiol Biomarkers Prev*. 2000; 9:895–903. [PubMed: 11008906]
16. Garner C. Upward bias in odds ratio estimates from genome-wide association studies. *Genet Epidemiol*. 2007; 31:288–95. [PubMed: 17266119]
17. Greenland S. Principles of multilevel modelling. *Int J Epidemiol*. 2000; 29:158–67. [PubMed: 10750618]
18. McCarthy MI, Zeggini E. Genome-wide association studies in type 2 diabetes. *Current Diab Rep*. 2009; 9:164–71.
19. Thomas DC, Clayton DG. Betting odds and genetic associations. *J Natl Cancer Inst*. 2004; 96:421–23. [PubMed: 15026459]
20. Wakefield J. A Bayes factors for genome-wide association studies: comparison with P -values. *Genet Epidemiol*. 2008; 33:79–86. [PubMed: 18642345]
21. Zollner S, Pritchard JK. Overcoming the winner’s curse: estimating penetrance parameters from case-control data. *Am J Hum Genet*. 2007; 80:605–15. [PubMed: 17357068]
22. Yu K, Chatterjee N, Wheeler W, Li Q, Wong S, Rothman N, Wacholder S. Flexible designs for following up positive findings. *Am J Hum Genet*. 2007; 81:540–51. [PubMed: 17701899]
23. Zhong H, Prentice RL. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics*. 2008; 9:621–34. [PubMed: 18310059]
24. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: applications to analysis of sequence data. *Am J Hum Genet*. 2008; 83:311–21. [PubMed: 18691683]
25. Morris CN. Parametric empirical Bayes inference: theory and applications. *J Am Stat Assoc*. 1983; 78:47–55.

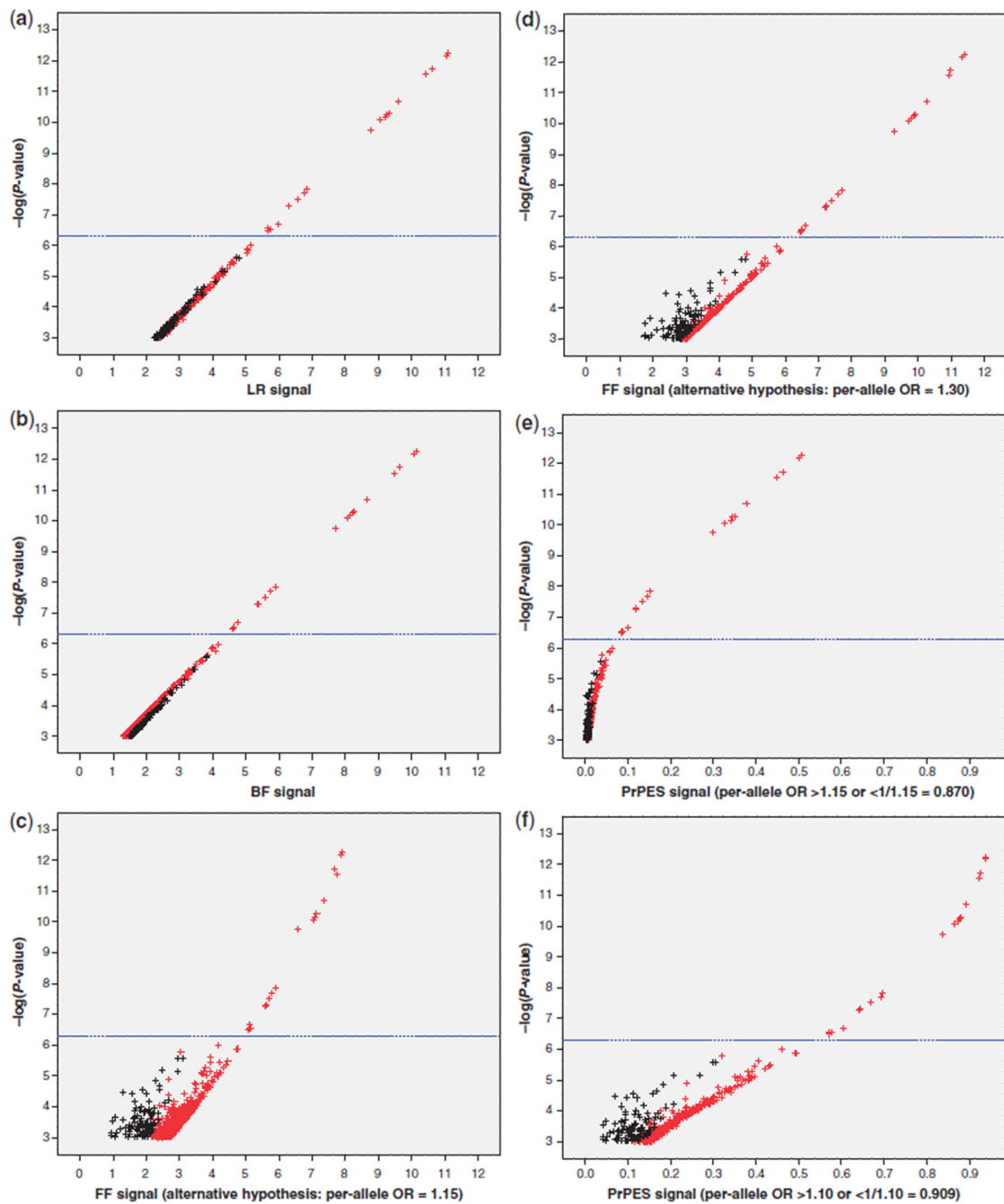


Figure 1.

Scatter plots showing the five types of empirical association signals calculated for the 651 SNPs with trend $P < 0.001$. The points are coloured according to the MAF among the controls [black, MAF $< 10\%$ ($n = 104$); red, MAF $> 10\%$ ($n = 547$)]. The horizontal reference line denotes a prior P -value threshold of 5×10^{-7} [$-\log_{10}(P\text{-value}) = 6.3$], derived by considering the false-positive report probability (association signals with $P < 5 \times 10^{-7}$ have been labelled as strong evidence of association; ref.²). (a) P -values vs LR signals [i.e. $\log_{10}(\text{LR})$]. (b) P -values vs BF signals [i.e. $\log_{10}(\text{BF})$]; BF obtained by considering a log-additive effect model, with specified prior distributions on the model parameters, and the null effect model]. (c and

d) *P*-values vs FF signals [i.e. $\log_{10}(\text{'FF'})$; 'FF' reflecting the ratio between post-data 'power' (for the alternative effect size given in each panel) and *P*-value]. **(e and f)** *P*-values vs PrPES signals [i.e. probabilities of pronounced effect size, defined as the per-allele OR, predicted towards the minor allele, above or below given limits; per-allele OR estimated by using (i) the log-additive effect model without specified prior distribution on the model parameters and (ii) a semi-Bayes adjustment procedure]. The rank-correlations (Spearman) equal 0.987, 0.969, 0.594, 0.836, 0.837 and 0.759 for the values shown in panels a, b, c, d, e and f, respectively.

Table 1

Statistical power at the significance level 5×10^{-7} for a candidate SNP in a GWA study with 2000 cases and 3000 unmatched controls, under three different disease models (single genetic marker effect), varying the expected risk allele frequency among the controls (Hardy–Weinberg equilibrium assumed)^a

Risk allele frequency	Statistical power under given disease model (OR)		
	Log-additive (1.3 ^b)	Dominant (1.3)	Recessive (1.3)
0.05	0.020	0.015	<0.001
0.10	0.170	0.095	<0.001
0.15	0.414	0.200	<0.001
0.20	0.625	0.277	<0.001
0.25	0.763	0.311	0.004
0.30	0.843	0.307	0.011
0.35	0.886	0.274	0.029
0.40	0.908	0.223	0.062
0.45	0.917	0.164	0.110
0.50	0.915	0.108	0.168

^aIt is assumed that the causal variant, or an SNP in perfect LD with it, will be typed. The results were obtained by using the freely available software program QUANTO (<http://hydra.usc.edu/GxE/>).

^bPer-allele odds ratio (OR).

Table 2Results for a few robustly replicating type 2 diabetes loci^a

SNP	OR ^b (95% CI)	Control MAF (%)	P	P-value	LR	BF	FF ^c	PrPES ^c	Signal rank	
									FF ^c	PrPES ^c
Robustly replicating SNPs										
rs864745	0.86 (0.79–0.93)	51	1.8E-4	179	182	203	113	245 ^d		
rs7961581	1.23 (1.13–1.35)	28	3.6E-6	30	27	30	27	24		
rs10946398	1.20 (1.10–1.31)	32	2.5E-5	65	57	68	51	37		
rs5015480	0.83 (0.76–0.90)	43	5.4E-6	32	41	34	24	40		
rs10811661	0.82 (0.73–0.92)	17	7.6E-4	528	557	488	508	376		
rs7901695	1.36 (1.25–1.48)	32	6.7E-13	2	2	2	2	2		
rs8050136	1.26 (1.16–1.37)	40	2.0E-8	12	12	12	12	12		
SNPs with notably different results—examples										
rs657317	1.66 (1.30–2.12)	2.2	3.8E-5	74	71	72	633	489		
rs12086219	0.62 (0.48–0.78)	3.9	6.9E-5	97	118	78	631	580		
rs10806665	1.42 (1.22–1.63)	7.7	2.7E-6	25	24	25	207	39 ^e		

^a Also provided are three examples of SNPs that were ranked notably different by the investigated association signal measures.^b Per-allele OR predicted towards the minor allele.^c The FF signal measure was calculated under the alternative OR=1.15 (analogous to OR=1/1.15=0.87); the PrPES signal measure was calculated for the corresponding PES magnitude (OR>1.15 or OR<0.87).^d Semi-Bayes adjusted OR (95% credibility interval²⁵): 0.93 (0.88–0.98).^e Semi-Bayes adjusted OR (95% credibility interval²⁵): 1.08 (1.01–1.16).