OPEN ACCESS Freely available online

PLoS one

Improving Power of Genome-Wide Association Studies with Weighted False Discovery Rate Control and Prioritized Subset Analysis

Wan-Yu Lin[1,2]*, Wen-Chung Lee[1,3]

1 Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan, **2** Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Alabama, United States of America, **3** Research Center for Genes, Environment and Human Health, National Taiwan University, Taipei, Taiwan

Abstract

The issue of large-scale testing has caught much attention with the advent of high-throughput technologies. In genomic studies, researchers are often confronted with a large number of tests. To make simultaneous inference for the many tests, the false discovery rate (FDR) control provides a practical balance between the number of true positives and the number of false positives. However, when few hypotheses are truly non-null, controlling the FDR may not provide additional advantages over controlling the family-wise error rate (e.g., the Bonferroni correction). To facilitate discoveries from a study, weighting tests according to prior information is a promising strategy. A 'weighted FDR control' (WEI) and a 'prioritized subset analysis' (PSA) have caught much attention. In this work, we compare the two weighting schemes with systematic simulation studies and demonstrate their use with a genome-wide association study (GWAS) on type 1 diabetes provided by the Wellcome Trust Case Control Consortium. The PSA and the WEI both can increase power when the prior is informative. With accurate and precise prioritization, the PSA can especially create substantial power improvements over the commonly-used whole-genome single-step FDR adjustment (i.e., the traditional un-weighted FDR control). When the prior is uninformative (true disease susceptibility regions are not prioritized), the power loss of the PSA and the WEI is almost negligible. However, a caution is that the *overall* FDR of the PSA can be slightly inflated if the prioritization is not accurate and precise. Our study highlights the merits of using information from mounting genetic studies, and provides insights to choose an appropriate weighting scheme to FDR control on GWAS.

Citation: Lin W-Y, Lee W-C (2012) Improving Power of Genome-Wide Association Studies with Weighted False Discovery Rate Control and Prioritized Subset Analysis. PLoS ONE 7(4): e33716. doi:10.1371/journal.pone.0033716

Editor: Dan E. Arking, Johns Hopkins University, United States of America

Received September 26, 2011; **Accepted** February 16, 2012; **Published** April 9, 2012

Funding: Funding for the WTCCC project was provided by the Wellcome Trust under award 076113. This study was partly supported by National Science Councils, Taiwan (NSC 99-2628-B-002-061-MY3). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: wlin@uab.edu

Introduction

The issue of large-scale testing has caught much attention with the advent of high-throughput technologies such as whole genome single-nucleotide polymorphism (SNP) arrays. In genome-wide association studies (GWAS), researchers are often confronted with a large number of SNPs. Two measures are commonly used to quantify the overall error rates when making simultaneous inference for the many SNPs. One is the family-wise error rate (FWER), defined as the probability of committing at least one type-I error from among a family of tests. Methods such as the Bonferroni correction and the Holm's step-down procedure [1] can be used to control the FWER. The other commonly used measure is the false discovery rate (FDR), defined as the expected ratio of the number of false rejections to the number of total rejections [2–6]. When the number of true null hypotheses (m_0) is smaller than the total number of hypotheses (m) (i.e., not all the null hypotheses are true), the FDR is smaller than or equal to the FWER [2]. Therefore, given a same nominal control level, controlling the FDR is less stringent than controlling the FWER. Controlling the FDR can provide a more practical balance between the number of true positives and the number of false positives. The FDR controlling has been widely applied to many gene expression data sets, in which the proportions of signal genes ($1-m_0/m$) are usually not small.

For some GWAS where the proportions of signal SNPs are extremely small ($m_0 \sim m$) [7], controlling the FDR provides no more benefits than controlling the FWER [8]. Fortunately, appropriately utilizing information from mounting genetic studies can improve this. If researchers have informative prior knowledge, weighting tests according to this prior information can substantially improve the power of a study [8,9]. There are two approaches to weight tests. One is the 'weighted FDR control' (WEI) [10]. The p values are weighted directly based on prior knowledge, and then the Benjamini and Hochberg's FDR controlling [2] or the Storey and Tibshirani's FDR controlling [4] is applied to the weighted p values. A study has shown a prominent benefit of using prior linkage results to weight the p values of association tests [9]. The second approach is the 'prioritized subset analysis' (PSA) [11], which has been applied to both GWAS [8] and gene expression data analyses [12]. A researcher first allocates all tests under study into two subsets,

based on his/her prior knowledge. A 'prioritized subset' comprises tests likely to be the true positives, and a 'non-prioritized subset' comprises the remaining tests. The FDR controlling is then applied to the two subsets, respectively.

Appropriately utilizing prior information is crucial to exploring signals, especially when the numbers of tests are going into millions, such as the scenario in GWAS. Both the WEI and the PSA have caught much attention for their advantage to facilitate discoveries from GWAS [8,9,11]. However, the comparison between them is not clear. In this work, we make a head-to-head comparison between these two approaches. We compare them with extensive simulations, and demonstrate their use on a real GWAS. Our work can provide insights to choose an appropriate strategy when making simultaneous inference for the many SNPs in GWAS.

Methods

Whole-genome Single-step FDR Adjustment (WGA)

The WGA is simply the traditional un-weighted FDR control for GWAS. Suppose that there are a total of m SNPs. Let $\{P_1,P_2,\cdots,P_m\}$ be the set of observed p values of the m SNPs. To control the FDR at a desired level, say q^*, the Benjamini and Hochberg's procedure [2] can be applied to this set of p values.

Prioritized Subset Analysis (PSA)

To perform the PSA and the following WEI, prior information is required to assign each SNP to be 'more likely a true positive' or 'more likely a true negative'. Suppose that we have prior information coming from our biological knowledge, or from findings of data other than that in the current study. Let $U_i = 1$ if the ith SNP is located in a chromosomal region supported by prior information and this SNP is thought to be more likely a true positive; $U_i = 0$ if this SNP is in a chromosomal region not supported by prior information and it is thought to be more likely a true negative. To perform a PSA, we first allocate all SNPs into two subsets: a 'prioritized subset' comprises SNPs likely to be the true positives ($U = 1$), and a 'on-prioritized subset' comprises the remaining SNPs ($U = 0$). The observed p values of the m SNPs are accordingly allocated into two subsets. One comprises the p values of the prioritized SNPs, and the other comprises the p values of the remaining non-prioritized SNPs. The Benjamini and Hochberg's FDR controlling [2] is then applied to these two subsets of p values, respectively.

Weighted False Discovery Rate Control Procedure (WEI)

There are two weighting schemes for the WEI: 'binary weighting' and 'general weighting' [10]. The 'general weighting' requires a researcher to assign a weight for each and every SNP specifically. To have a parallel comparison between the WEI and the PSA, we here only consider the 'binary weighting' for the WEI. In 'binary weighting', SNPs thought to be more likely true positives ($U = 1$) are all assigned a same weight (w_1), and SNPs thought to be more likely true negatives ($U = 0$) are all assigned another weight (w_0).

Let W_i be the weight assigned to the ith SNP, $i = 1, \ldots, m$. In the binary weighting scheme, W_i is either w_1 or w_0. The p values are weighted according to $P_i^* = P_i/W_i$, where P_i^* is the weighted p value of the ith SNP. The Benjamini and Hochberg's FDR controlling [2] is then applied to the set of weighted p values $\{P_1^*,P_2^*,\cdots,P_m^*\}$. To maintain the FDR at a desired level, the set of weights $\{W_1,W_2,\cdots,W_m\}$ must meet a requirement: $\bar{W} = \sum_{i=1}^{m} W_i/m = 1$ [10]. For the weights in the binary weighting scheme, a researcher can first decide a w_1 (or w_0), and work out the other one, w_0 (or w_1), using the constraint $\bar{W} = 1$. An alternative is to choose a ratio of the two weights, $r = w_1/w_0 \geq 1$, and then obtain w_0 with the constraint:

$$1 = \bar{W} = w_0 \times (1-\bar{U}) + w_1 \times \bar{U} = w_0 \times (1-\bar{U}) + r w_0 \times \bar{U},$$

where $\bar{U} = \sum_{i=1}^{m} U_i/m$. Therefore, $w_0 = 1/[1+(r-1)\bar{U}]$ and $w_1 = r/[1+(r-1)\bar{U}]$.

The choice of r reflects the degree of confidence a researcher has toward the prior, which is subjective and is specified by the researcher. (Note that the PSA does not require this parameter, because the PSA simply allocates all SNPs into two subsets without specifying any explicit weight.) If the researcher is confident of the prior information, r can be specified larger. If not, r should be specified smaller. When $r = 1$, $w_1 = w_0$, the WEI reduces to the WGA.

Simulations

We performed simulations to compare the power and the ability to control the FDR of the WGA, the PSA, and the WEI. To provide a practical evaluation on these methods when analyzing GWAS, we followed Li et al. [11] to first simulate GWAS data with similar linkage disequilibrium (LD) patterns as the HapMap data [13], and then analyzed the simulated GWAS data sets with the WGA, the PSA, and the WEI. For the WEI, we followed Genovese et al. [10] to specify r at 2, 5, or 10.

Simulation program. We used a rapid whole-genome simulation program, the GWAsimulator [14] (http://biostat.mc.vanderbilt.edu/wiki/Main/GWAsimulator), to generate GWAS data sets. The GWAsimulator [14] implements a rapid moving-window algorithm [15] to simulate whole genome case-control or population samples. It faithfully generates SNP genotypes that follow the local LD patterns of the input data. Following Li et al. [11], we used the phased data of HapMap 60 CEU (CEPH samples with ancestry from northern and western Europe) founder subjects as the input data. The total number of SNPs in the input data is 314,174, after merging the Illumina Sentrix HumanHap300 BeadChips (317,503 SNPs) and the HapMap phased data [11,14].

Setting of the disease model. Following Li et al. [11], we let six SNPs be the disease variants. Among the six variants, three have small effects (genotypic relative risk or GRR = 1.34) and the others have relatively large effects (GRR = 1.57). We randomly chose SNPs with minor allele frequencies (MAFs) of 0.25, 0.36, and 0.33 from chromosomes 6, 10, and 5 respectively, as the three small-effect SNPs (Locus 1–3). The three large-effect SNPs (Locus 4–6) with MAFs of 0.43, 0.31, and 0.30 were randomly picked from chromosomes 3, 11, and 4 respectively. This MAF setting mimics the reported risk loci of type 2 diabetes [16], in which the minor alleles were treated as risk alleles. Given the genotypes of the six disease loci, the probability of being affected is $\Pr(affected|genotype) = \left[1+\exp(-\beta_0 - \sum_{j=1}^{6}\beta_j g_j)\right]^{-1}$, where $g_j \in \{0,1,2\}$ is the number of risk alleles at disease locus j, β_0 was chosen to lead to 5% of the population disease prevalence, and β_j's were chosen to meet the specification for the six GRRs.

We simulated 15,000 replicate data sets. In each replication, genotypes of 314,174 whole-genome SNPs were generated for each of 500 unrelated cases and 500 unrelated controls. SNPs with Hardy-Weinberg exact P value$<10^{-3}$ in the control group were excluded. P values were obtained using the one-degree-of-freedom chi-square test to compare the allele frequencies in cases and controls. The FDR level was to be controlled at 5%.

We followed Li et al. [11] to define the prioritized subsets according to the combinations of three factors: (Factor A) the number of prioritized regions (6, 14, 22); (Factor B) the size of each prioritized region (2 Mb, 20 Mb); (Factor C) the disease loci to be prioritized: (i) no disease loci; (ii) Locus 6; (iii) Loci 1 and 6; (iv) Loci 1, 2, 5, and 6; (v) Loci 4, 5, and 6; (vi) Loci 1, 4, 5, and 6; (vii) all except Locus 3; (viii) all loci. Factors A and B are directly related to the precision of a prioritization process, while Factor C is related to its accuracy.

Results

Simulation Results

FDR evaluations. Figures 1 and 2 present the FDR (the mean ratio of the number of false rejections to the number of total rejections, based on the 15,000 replications) of the WGA, the PSA, and the WEI ($r = 2, 5, 10$) when the prioritized region sizes were 2 Mb and 20 Mb, respectively. Note that the FDR of the PSA was the *overall* FDR, obtained by the mean ratio of the number of *total* false rejections *from the two subsets* to the number of *total* rejections *from the two subsets*. The WGA and the WEI had very similar FDRs, which were under control for all scenarios. For the PSA, accurate (true disease loci were prioritized) and precise (prioritized regions were narrower) prioritization led to lower *overall* FDR. Comparing the eight scenarios when six 2-Mb regions were prioritized (Figure 1), the more accurate the prioritization (more true disease loci were prioritized), the lower the *overall* FDR. However, within a same scenario (especially for Scenarios (iv)–(viii)), the *overall* FDRs were not as low when more regions unrelated to the disease were prioritized (14 or 22 regions compared to 6 regions). Comparing Figure 2 (where six 20-Mb regions were prioritized) with Figure 1 (where six 2-Mb regions were prioritized), the *overall* FDRs did not remain as low due to the decreasing precision of the prioritization (20 Mb compared to 2 Mb). The ability of the PSA to control the *overall* FDR depends on the accuracy and precision of the prioritization. When none of the true disease loci was prioritized (Scenario (i) of Figures 1 and 2), the *overall* FDR of the PSA was inflated to 5.6%. When only one or two disease loci were prioritized with higher precision (Scenarios (ii) & (iii) of Figure 1), the *overall* FDR of the PSA was 'almost' under control (still a slight inflation on the FDR). However, when the one or two disease loci were prioritized with lower precision (Scenarios (ii) & (iii) of Figure 2), the *overall* FDR of the PSA was inflated to 5.3%.

We see that the *overall* FDR of the PSA can be inflated to 5.6%. To have a fair evaluation on the power, we deliberately set the FDR control level at 4.4% (because the ratio of 5% to 5.6% is approximately equal to the ratio of 4.4% to 5%) for the PSA, under Scenarios (i)–(iii) for 2 Mb prioritized region size and under Scenarios (i)–(v) for 20 Mb size. After this adjustment, the *overall* FDRs of the PSA method were not larger than 5% (see our Supporting Information S1).

Power evaluations. Power of detecting a disease locus was defined as the proportion of 'successful detections' of that disease locus among all the 15,000 replications, in which a 'successful detection' was defined as 'declaration of significance for at least one SNP within one Mb from the disease locus' (following the definition by Li et al. [11]). Figures 3 and 4 present the power comparisons between the WGA, the PSA, and the WEI ($r = 2, 5, 10$) when six 2-Mb regions and when six 20-Mb regions were prioritized, respectively. Our Supporting Information S1 shows the results when 14 or 22 regions were prioritized, each region with a size of 2 Mb or 20 Mb. Note that the power was compared (Figures 3, 4, and Supporting Information S1) with the adjustment of FDR for the PSA (that is, the FDR control level was deliberately set at 4.4% for the PSA, under Scenarios (i)–(iii) for 2 Mb prioritized region size and under Scenarios (i)–(v) for 20 Mb size.).

The accuracy of a prioritization process directly affects the relative power performance of the three methods. When the regions encompassing the true disease loci were prioritized, the most powerful method was the PSA, then was the WEI ($r = 10$). The WEI with a larger r was more powerful than the WEI with a smaller r (see Scenarios (ii)–(viii) of Figures 3 & 4, where at least one disease locus was prioritized). When the regions encompassing the true disease loci were *not* prioritized, the power loss of the PSA and the WEI was almost negligible. The WEI with a larger r suffered from more power loss than the WEI with a smaller r (see Scenario (i) of the last figure in the Supporting Information S1, where no disease loci were prioritized).

Furthermore, the precision of a prioritization process also influences the power of the PSA and the WEI. Regarding the size of each prioritized region (2 Mb or 20 Mb) and the number of prioritized regions (6, 14, or 22), the power improvement of the PSA and the WEI was not as prominent when the prioritization was not as precise, i.e., a wider region around a true disease locus was prioritized (20 Mb compared to 2 Mb), or more regions unrelated to the disease were prioritized (22 or 14 regions compared to 6 regions).

Application to the Wellcome Trust Case Control Consortium (WTCCC) Data

We further demonstrate the WGA, the PSA, and the WEI with a GWAS on type 1 diabetes (T1D). The data set was provided by the Wellcome Trust Case Control Consortium (WTCCC) [17] that included 2,000 T1D cases and 3,000 controls. Subjects were living within England, Scotland, and Wales ('Great Britain'). The vast majority had self-identified themselves as white Europeans [17]. The control subjects were from 1958 British Birth Cohort (1,500 subjects) and UK Blood Services sample (1,500 subjects). After excluding subjects identified as having recent non-European ancestry, there were 1,963 T1D cases and 2,938 controls [17]. Subjects were genotyped using the Affymetrix GeneChip 500 K arrays comprising 500,568 SNPs. According to the WTCCC criteria [17], 459,653 SNPs passed the quality control filters. We further removed 578 SNPs with poor clustering and retained 459,075 SNPs.

WGA. For each SNP, we used the p value obtained from the genotypic test. Controlling the FDR at 5%, 12 independent association signals were declared to be significant with the WGA (an association signal was identified given more than a single significant SNP within 2 Mb). Figure 5 and Table 1 show the 12 signals that can be mapped to 12 genes. Among them, five were declared to be significant when the Bonferroni correction was used to control the FWER at 5%.

Prioritization process. To perform the PSA and the WEI, prior information is required to assign each SNP to be 'more likely a true positive' or 'more likely a true negative'. We collected the information by searching for publications with 'gene', 'association', and 'type 1 diabetes' in their titles. PubMed shows 89 publications meeting this searching criterion. Among them, we used the studies independent of the WTCCC project and published *prior to* the WTCCC publication [17]. Table 2 lists the genes that are previously reported to be associated with T1D. We obtained the physical position of each gene (listed in Table 2) from the Gene Location website (http://genecards.weizmann.ac.il/geneloc/index.shtml). SNPs within 1 Mb from each gene were prioritized (so the size of each prioritized region was 2 Mb). When prioritizing SNPs according to prior information, there is no stringent criterion for the sizes of prioritized regions (we will discuss this in the

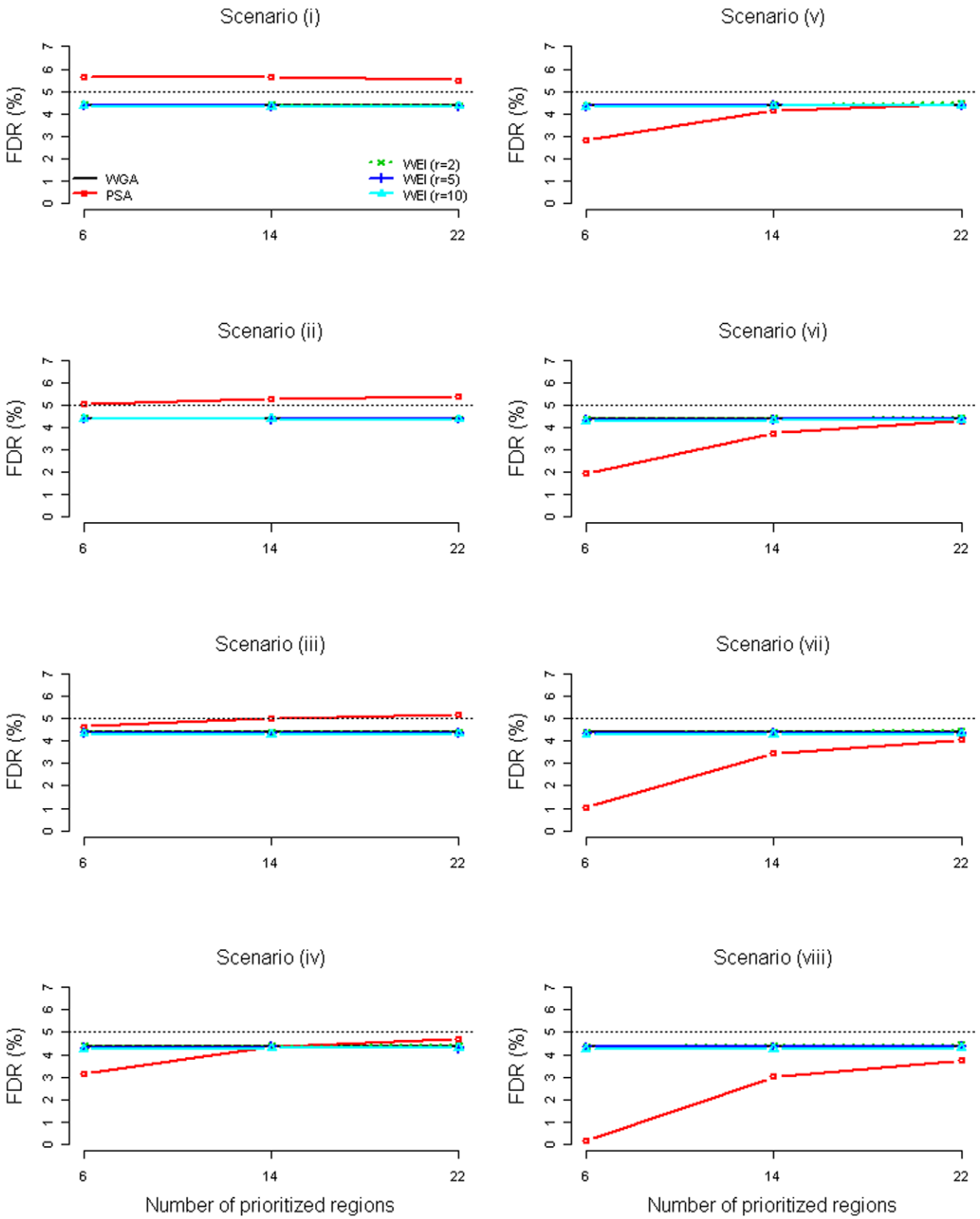

Figure 1. FDR of the WGA, the PSA, and the WEI ($r = 2, 5, 10$) when the prioritized region sizes were 2 Mb (without adjustment to the PSA).
doi:10.1371/journal.pone.0033716.g001

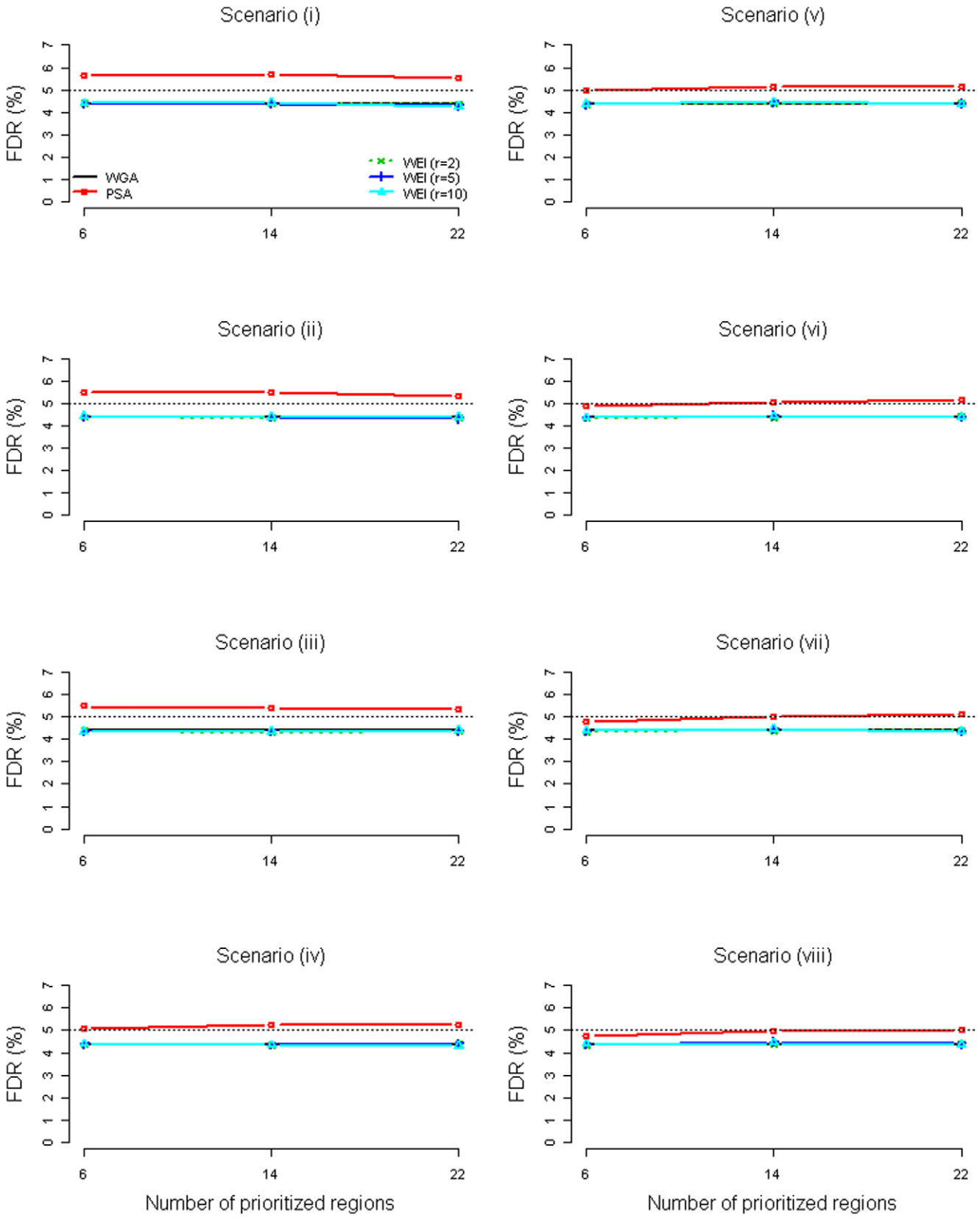

Figure 2. FDR of the WGA, the PSA, and the WEI (r = 2, 5, 10) when the prioritized region sizes were 20 Mb (without adjustment to the PSA).
doi:10.1371/journal.pone.0033716.g002

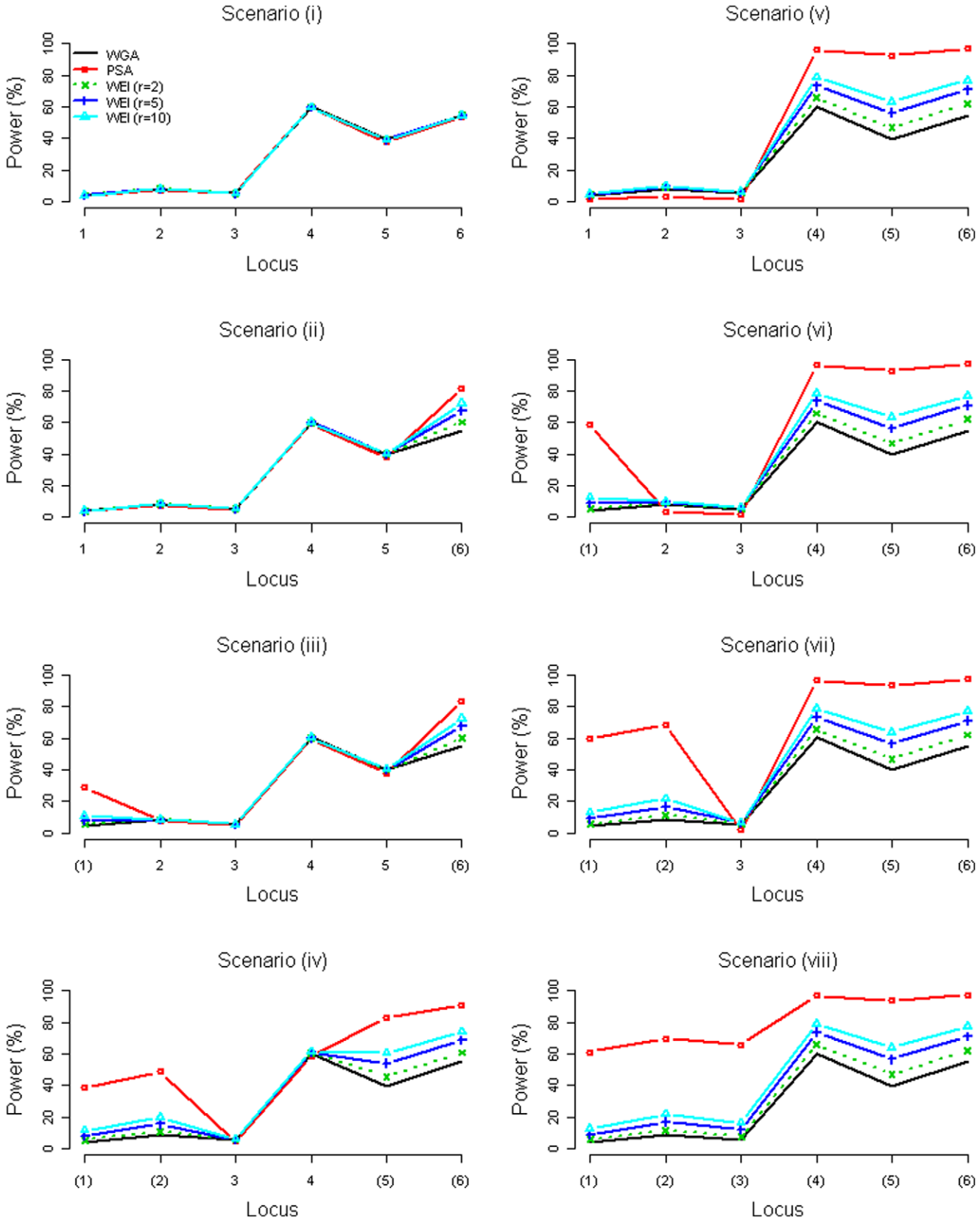

Figure 3. Power comparison between the WGA, the PSA, and the WEI (r=2, 5, 10) when six 2-Mb regions were prioritized (with adjustment to the PSA). A locus with parentheses indicates that the disease locus was included in the prioritized subset.
doi:10.1371/journal.pone.0033716.g003

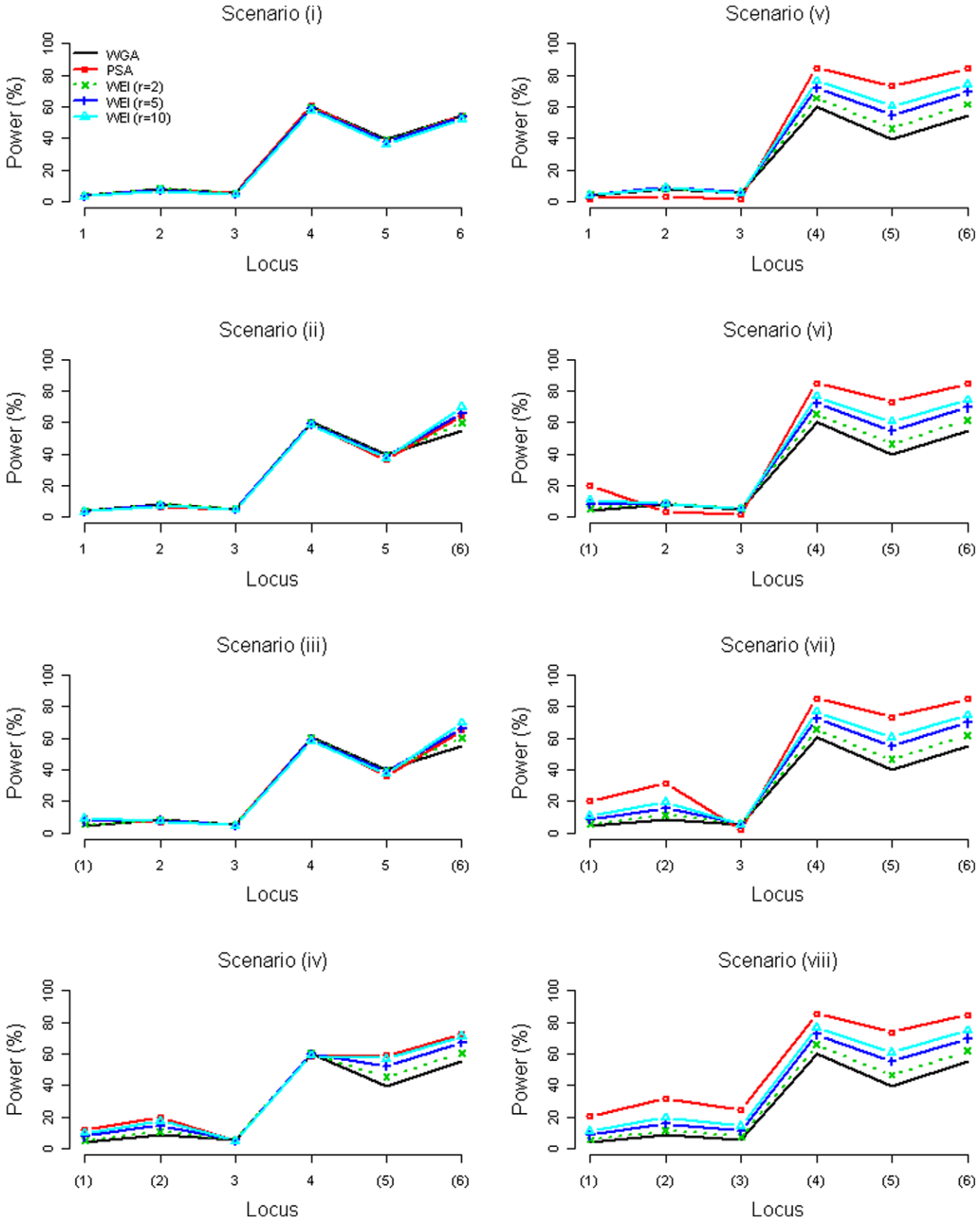

Figure 4. Power comparison between the WGA, the PSA, and the WEI ($r = 2, 5, 10$) when six 20-Mb regions were prioritized (with adjustment to the PSA). A locus with parentheses indicates that the disease locus was included in the prioritized subset.
doi:10.1371/journal.pone.0033716.g004

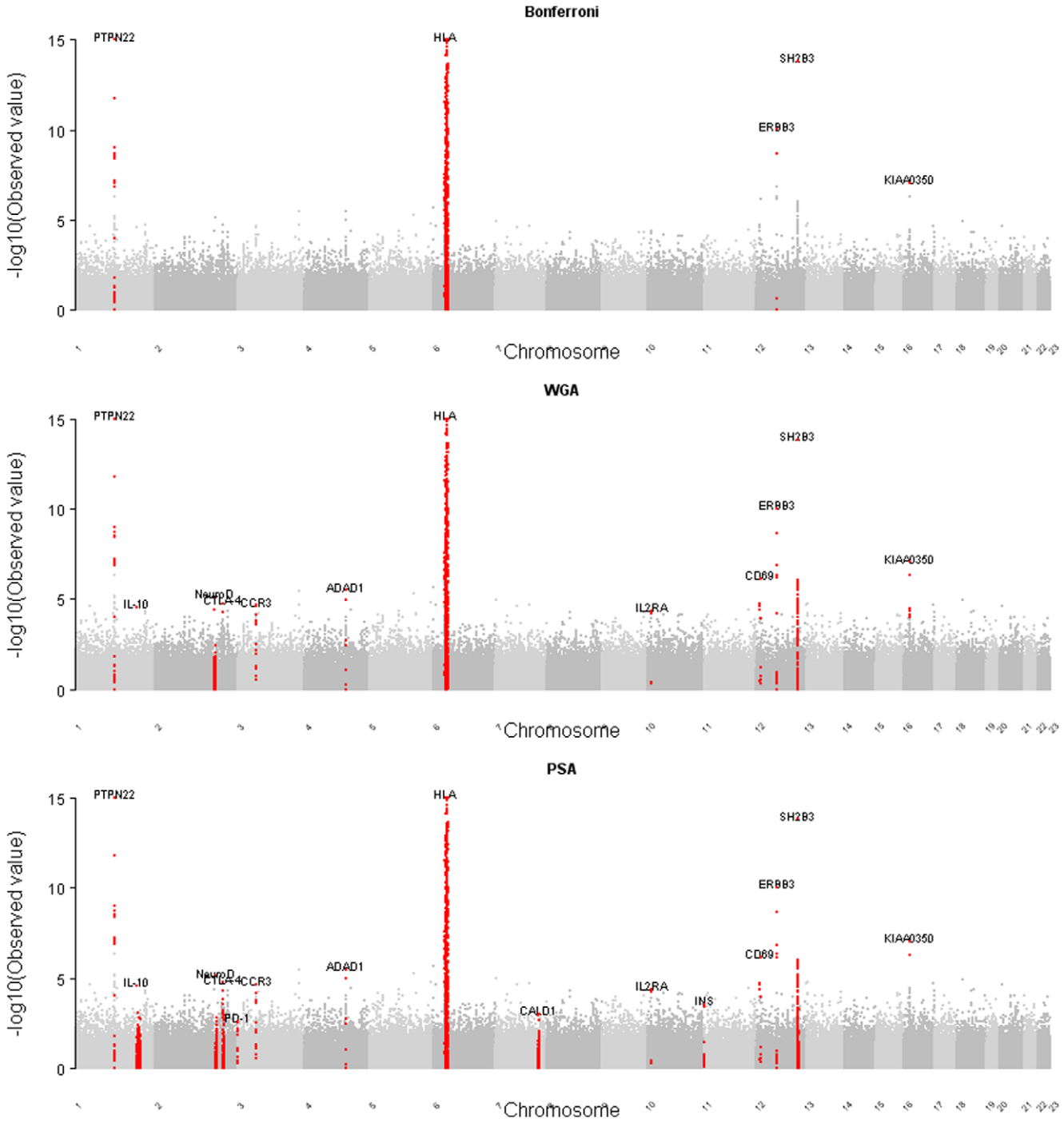

Figure 5. Manhattan plot for the T1D data set. The *x*-axis lists the chromosome numbers, and the *y*-axis presents $-\log_{10}(p\ value)$. The red points label the significant genes identified by each method (from top to bottom: the Bonferroni correction to control the FWER at 5%, the WGA to control the FDR at 5%, and the PSA to control the FDR at 5%). This figure was plotted with the R package 'gap' [31].
doi:10.1371/journal.pone.0033716.g005

Discussion). However, in general, a wider region should be prioritized for a linkage peak than that for an association signal, because linkage is a coarse mapping while association is a fine mapping. In this data analysis, our prior information came from previous association studies on T1D, and therefore we chose a moderate prioritization size for each region −2 Mb. Totally, we prioritized 6,914 SNPs and left the remaining 452,161 SNPs in the non-prioritized subset. A list of the prioritized SNPs is available upon request.

PSA. To perform a PSA, the genotypic p values of the 459,075 SNPs were accordingly allocated into two subsets. The FDR was to be controlled at 5%. The Benjamini and Hochberg's FDR controlling [2] was applied to the two subsets of p values, respectively. Finally, 15 genes were declared to be significant (see Figure 5 and Table 1). Among them, nine genes came from the

Table 1. Results of the T1D data set.

Gene[1]	Chromosome	Bonferroni[2]	WGA, WEI ($r=2$)	WEI ($r=5$, 10)	PSA	WTCCC SNP[3]	Genotypic P value[4]	Supported by later studies[5]
HLA*	6p21	V	V	V	V	rs9272346	5.47×10^{-134}	[32–34]
PTPN22*	1p13	V	V	V	V	rs6679677	5.43×10^{-26}	[22,35–49]
SH2B3	12q24	V	V	V	V	rs17696736	1.51×10^{-14}	[50]
ERBB3	12q13	V	V	V	V	rs11171739	9.71×10^{-11}	[51]
KIAA0350	16p13	V	V	V	V	rs12708716	4.92×10^{-7}	[52,53]
CD69	12p13		V	V	V	rs11052552	7.24×10^{-7}	[54]
ADAD1	4q27		V	V	V	rs17388568	3.27×10^{-6}	[55–57]
NeuroD*	2q32		V	V	V	rs10206282	7.89×10^{-6}	[58,59]
CTLA4*	2q33		V	V	V	rs231726	1.78×10^{-5}	[19,60–63]
CCR3	3p21		V	V	V	rs2157057	2.20×10^{-5}	[64]
IL-10*	1q31–q32		V	V	V	rs12061474	2.66×10^{-5}	[65–70]
IL2RA*	10p15–p14		V	V	V	rs2104286	4.32×10^{-5}	[71–76]
INS*	11p15			V	V	rs6578252	2.82×10^{-4}	[19–22]
CALD1*	7q33				V	rs2250603	9.53×10^{-4}	[23,24]
PD-1*	2q37				V	rs10192057	2.59×10^{-3}	[25,26]
Number of significant genes		5	12	13	15			

1 *: being allocated in the prioritized subset.
[2]V: significant.
[3]The WTCCC SNP showing the strongest association evidence in that region.
[4]The p value of the genotypic test of the WTCCC SNP showing the strongest association evidence in that region.
[5]That association signal is supported by later studies that have *NOT* been selected as our prior information (in Table 2).
doi:10.1371/journal.pone.0033716.t001

prioritized subset, including *HLA*, *PTPN22*, *NeuroD*, *CTLA4*, *IL-10*, *IL2RA*, *INS*, *CALD1*, and *PD-1*. The last three were not identified as significant genes with the WGA method. Being allocated into the prioritized subset, they (*INS*, *CALD1*, and *PD-1*) had a chance to be identified as significant genes. We further evaluated whether the FDR was well controlled in the prioritized subset. We permuted the phenotypes 10^4 times and obtained 10^4 null p values for each of the 6,914 SNPs. Then we estimated the number of false positives by counting the number of null p values more extreme (i.e., smaller) than 2.59×10^{-3}, the largest P value of the association signals in the prioritized subset. In this way, we obtained an estimated permutation-based FDR [18] at 3.64%, still less than the FDR control level of 5%. In fact, several studies published later than the WTCCC paper [17] also supported the association of these three genes with T1D (*INS* [19–22], *CALD1* [23,24], and *PD-1* [25,26]).

WEI. To perform a WEI, in addition to the prior information used for the PSA, a ratio of the two weights (r) was specified at 2, 5, and 10, respectively. In this T1D study, the proportion of SNPs thought to be more likely true positives among all the SNPs was $\bar{U}=\sum_{i=1}^{m}U_i/m=6914/459075=0.015$. With a specified r, the weight given to SNPs thought to be more likely true positives was $w_1=r/[1+(r-1)\bar{U}]$, and that given to SNPs thought to be more likely true negatives was $w_0=1/[1+(r-1)\bar{U}]$. The p values were then weighted according to $P_i^*=P_i/W_i$, where P_i and P_i^* were respectively the original and weighted p values of the ith SNP, and $W_i\in\{w_0,w_1\}$. We used the Benjamini and Hochberg's method [2] to control the FDR at 5%. When r was specified at 2, the WEI identified the same 12 genes as the WGA. When r was specified at 5 or 10, the WEI identified one more gene – *INS* (see Table 1).

Discussion

The PSA and the WEI both require prior knowledge to boost the power of detecting signals. Prior knowledge can be collected from previous independent studies that were not based on the same data of the current working study. It should be searched *before* seeing the analysis results of the individual tests in the current study. Both the PSA and the WEI have caught much attention in the era of high-throughput genomics, for their improvement on the FDR control. With the advancement of biological technologies, we now can obtain substantial genomic data with decreasing costs. To facilitate discoveries from more and more hypothesis tests, clarifying the merits and limitations of the PSA and the WEI is important.

The PSA and the WEI both can increase power when the prior is informative. As shown in our simulation, the PSA can especially create substantial power improvements given accurate and precise prioritization. When researchers fail to prioritize some true disease loci, the power loss of the PSA and the WEI is almost negligible. Like the WGA, the WEI has a solid theoretical background [10] and a good ability to control the FDR at the desired level (Figures 1 and 2).

Although the PSA can increase much power when the prior is informative, its *overall* FDR can be slightly inflated if the prioritization is not accurate and precise. In our simulation, the *overall* FDR of the PSA was obtained by the mean ratio of the number of *total* false rejections *from the two subsets* to the number of *total* rejections *from the two subsets*, based on 15,000 replications. Because the Benjamini and Hochberg's FDR controlling [2] is applied to the prioritized and the non-prioritized subsets *respectively*, the *overall* FDR of the PSA is not guaranteed to be controlled at the desired level even when the numbers of SNPs in the two subsets

Table 2. Prior knowledge: genes that are previously reported to be associated with type 1 diabetes.

Gene	Chromosome	Start base pair[1]	End base pair[1]	Publications prior to the WTCCC paper
PTPN22	1	114356433	114414381	[77–79]
IL-10	1	206940947	206945839	[80]
IL1R1	2	102681004	102744178	[81]
NeuroD	2	182537815	182545603	[82]
CD28	2	204571198	204738683	[83]
CTLA4	2	204732509	204738683	[83–90]
SLC11A1[2]	2	219246752	219261617	[91,92]
PD-1	2	242792033	242801060	[93]
DBP	4	72607410	72669758	[94]
MIC-A	6	31367561	31433586	[95]
HLA-DRB1	6	32546546	32557625	[96,97]
HLA-DQA1	6	32595956	32714992	[97]
HLA-DQB1	6	32627244	32731330	[97]
SUMO4	6	149721495	149722182	[98]
MTH1	7	2281857	2291004	[99]
CALD1	7	134429003	134655480	[100]
IL2RA	10	6052652	6104333	[101]
INS	11	2181009	2182571	[102,103]
IL-18	11	112013974	112034840	[85,104]
VDR	12	48235320	48336831	[105–107]
OAS	12	113344582	113369991	[108]
HSD11B2	16	67465036	67471456	[109]
ICAM-1	19	10381517	10397291	[110,111]

[1]The physical positions were obtained from the Gene Location website (http://genecards.weizmann.ac.il/geneloc/index.shtml).
[2]Other aliases: LSH, NRAMP, NRAMP1.
doi:10.1371/journal.pone.0033716.t002

are both large enough. For example, when 22 20-Mb regions (not precise) were prioritized, the *overall* FDRs of the PSA were inflated under Scenarios (i)–(iv) (not accurate) despite ~49,000 SNPs in the prioritized subset and ~265,000 SNPs in the non-prioritized subset. This is because when the prioritization is not accurate and precise, the number of true positives from the prioritized subset can be small (or the number of false positives can be large). This can inflate the *overall* FDR of the PSA.

There is no stringent criterion for the sizes of prioritized regions, when prioritizing SNPs according to prior information. Given a linkage peak or an association signal based on prior studies, prioritizing a narrow region may miss the true disease loci and lose the accuracy, while prioritizing a wide region may lose the precision. It is not easy to conclude how wide a region should be prioritized. Although the determination for a region size is somewhat *ad hoc*, a general principle is to estimate the permutation-based FDR [18] after performing the PSA. This can empirically evaluate whether the FDR within each subset is well controlled.

The WEI can be equipped with 'general weighting scheme', although we only evaluated the 'binary weighting scheme' for its parallel comparison with the PSA. For the general weighting scheme, each test is assigned a specific weight, not only either w_1 or w_0. In this way, the WEI is more flexible than the PSA in the sense that the weights can be assigned in a continuous scale. To mimic this flexibility, the PSA can extend its original concept to allocate SNPs into more than two subsets. However, this will inevitably increase the possibility of unsatisfactory FDR control.

In addition to conventional GWAS, weighting tests can provide insights to rare variant detection. In the past several years, GWAS have identified hundreds of common genetic variants (minor allele frequency (MAF)>5%) for complex human diseases [27]. However, these common variants can only explain a small proportion of heritability. The field of genetic epidemiology is shifting toward the study of low-frequency (MAF 1%–5%) and rare variants (MAF<1%), which are thought to have larger effect sizes than common variants [28]. Unfortunately, rare variants are difficult to detect due to their low frequencies. Recently, a weighted-Holm procedure was shown to substantially improve the power of detecting rare variants with large genetic effects [29]. Furthermore, a study has shown that low-frequency variants can be identified by up-weighting SNPs with lower MAFs and then performing the FDR control [30]. Appropriately weighting genetic variants according to their MAFs can facilitate the detection of rare variants. Applying the PSA and the WEI to this topic deserves further investigation.

Supporting Information

Supporting Information S1 FDR of the WGA, the PSA, and the WEI ($r = 2, 5, 10$) when the prioritized region sizes were 2 Mb and 20 Mb (with adjustment to the PSA), respectively; power comparison between the WGA, the PSA, and the WEI ($r = 2, 5, 10$) when 14 2-Mb, 14 20-Mb, 22 2-Mb, and 22 20-Mb regions were prioritized (with adjustment to the PSA), respectively.
(DOC)

Acknowledgments

We thank the Academic Editor and the anonymous reviewers for their constructive comments. We also thank Drs. Chun Li and Mingyao Li for making their simulation program GWAsimulator available to the scientific community. This study makes use of data provided by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from http://www.wtccc.org.uk/info/participants.shtml.

Author Contributions

Conceived and designed the experiments: WYL WCL. Performed the experiments: WYL WCL. Analyzed the data: WYL. Wrote the paper: WYL WCL.

References

1. Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Statist 6: 65–70.
2. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B 57: 289–300.
3. Storey JD (2002) A direct approach to false discovery rates. J R Stat Soc B 64: 479–498.
4. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100: 9440–9445.
5. Strimmer K (2008) A unified approach to false discovery rate estimation. BMC Bioinformatics 9: 303.
6. Strimmer K (2008) fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics 24: 1461–1462.
7. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, et al. (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308: 385–389.
8. Lin W-Y, Lee W-C (2010) Incorporating prior knowledge to facilitate discoveries in a genome-wide association study on age-related macular degeneration. BMC Research Notes 3: 26.
9. Roeder K, Bacanu SA, Wasserman L, Devlin B (2006) Using linkage genome scans to improve power of association in genome scans. Am J Hum Genet 78: 243–252.
10. Genovese C, Roeder K, Wasserman L (2006) False discovery control with P-value weighting. Biometrika 93: 509–524.
11. Li C, Li M, Lange EM, Watanabe RM (2008) Prioritized subset analysis: improving power in genome-wide association studies. Hum Hered 65: 129–141.
12. Lin W-Y, Lee W-C (2011) Floating prioritized subset analysis: a powerful method to detect differentially expressed genes. Computational Statistics and Data Analysis 55: 903–913.
13. The International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437: 1299–1320.
14. Li C, Li M (2008) GWAsimulator: a rapid whole-genome simulation program. Bioinformatics 24: 140–142.
15. Durrant C, Zondervan KT, Cardon LR, Hunt S, Deloukas P, et al. (2004) Linkage disequilibrium mapping via cladistic analysis of single-nucleotide polymorphism haplotypes. Am J Hum Genet 75: 35–43.
16. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, et al. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445: 881–885.
17. WTCCC (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.
18. Xie Y, Pan W, Khodursky AB (2005) A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. Bioinformatics 21: 4280–4288.
19. Bjornvold M, Undlien DE, Joner G, Dahl-Jorgensen K, Njolstad PR, et al. (2008) Joint effects of HLA, INS, PTPN22 and CTLA4 genes on the risk of type 1 diabetes. Diabetologia 51: 589–596.
20. Fendler W, Klich I, Cieslik-Heinrich A, Wyka K, Szadkowska A, et al. (2011) Increased risk of type 1 diabetes in Polish children - association with INS-IGF2 5′VNTR and lack of association with HLA haplotype. Endokrynol Pol 62: 436–442.
21. Liu M, Hodish I, Haataja L, Lara-Lemus R, Rajpal G, et al. (2010) Proinsulin misfolding and diabetes: mutant INS gene-induced diabetes of youth. Trends Endocrinol Metab 21: 652–659.
22. Stene LC, Ronningen KS, Undlien DE, Joner G (2011) Does the relative risk for type 1 diabetes conferred by HLA-DQ, INS, and PTPN22 polymorphisms vary with maternal age, birth weight, or cesarean section? Pediatr Diabetes 12: 91–94.
23. Millioni R, Iori E, Lenzini L, Puricelli L, Caroccia B, et al. (2011) Caldesmon over-expression in type 1 diabetic nephropathy. J Diabetes Complications 25: 114–121.
24. Millioni R, Iori E, Puricelli L, Arrigoni G, Vedovato M, et al. (2008) Abnormal cytoskeletal protein expression in cultured skin fibroblasts from type 1 diabetes mellitus patients with nephropathy: A proteomic approach. Proteomics Clin Appl 2: 492–503.
25. Momin S, Flores S, Angel BB, Codner DE, Carrasco PE, et al. (2009) Interactions between programmed death 1 (PD-1) and cytotoxic T lymphocyte antigen 4 (CTLA-4) gene polymorphisms in type 1 diabetes. Diabetes Res Clin Pract 83: 289–294.
26. Ni R, Ihara K, Miyako K, Kuromaru R, Inuo M, et al. (2007) PD-1 gene haplotype is associated with the development of type 1 diabetes mellitus in Japanese children. Hum Genet 121: 223–232.
27. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747–753.
28. Zeggini E (2011) Next-generation association studies for complex traits. Nat Genet 43: 287–288.
29. Dalmasso C, Genin E, Tregouet DA (2008) A weighted-Holm procedure accounting for allele frequencies in genomewide association studies. Genetics 180: 697–702.
30. Xing C, Cohen JC, Boerwinkle E (2010) A weighted false discovery rate control procedure reveals alleles at FOXA2 that influence fasting glucose levels. Am J Hum Genet 86: 440–446.
31. Zhao JH (2007) gap: Genetic Analysis Package. Journal of Statistical Software 23.
32. Husain Z, Kelly MA, Eisenbarth GS, Pugliese A, Awdeh ZL, et al. (2008) The MHC type 1 diabetes susceptibility gene is centromeric to HLA-DQB1. J Autoimmun 30: 266–272.
33. Jung MH, Yu J, Shin CH, Suh BK, Yang SW, et al. (2009) Association of cytotoxic T lymphocyte antigen-4 gene polymorphisms and HLA class II alleles with the development of type 1 diabetes in Korean children and adolescents. J Korean Med Sci 24: 1004–1009.
34. Rohana AG, Loh KC, Tin SK, Soh CH, Nazaimoon WM, et al. (2011) HLA-DQ A1, -DQB1 and -DRB1 gene polymorphism–in Malay type 1 diabetes mellitus patients and their use for risk prediction. Med J Malaysia 66: 133–137.
35. Aarnisalo J, Treszl A, Svec P, Marttila J, Oling V, et al. (2008) Reduced CD4+T cell activation in children with type 1 diabetes carrying the PTPN22/Lyp 620Trp variant. J Autoimmun 31: 13–21.
36. Barone B, Dantas JR, Almeida MH, Anna-Gomes BS, Bencke-Gongalves MD, et al. (2011) Pancreatic autoantibodies, HLA DR and PTPN22 polymorphisms in first degree relatives of patients with type 1 diabetes and multiethnic background. Exp Clin Endocrinol Diabetes 119: 618–620.
37. Dieude P, Teixeira VH, Pierlot C, Cornelis F, Petit-Teixeira E (2008) Testing for linkage and association with rheumatoid arthritis a ptpn22 promoter polymorphism reported to be associated and linked with type 1 diabetes in the Caucasian population. Ann Rheum Dis 67: 900–901.
38. Douroudis K, Kisand K, Nemvalts V, Rajasalu T, Uibo R (2010) Allelic variants in the PHTF1-PTPN22, C12orf30 and CD226 regions as candidate susceptibility factors for the type 1 diabetes in the Estonian population. BMC Med Genet 11: 11.
39. Fichna M, Zurawek M, Januszkiewicz-Lewandowska D, Fichna P, Nowak J (2010) PTPN22, PDCD1 and CYP27B1 polymorphisms and susceptibility to type 1 diabetes in Polish patients. Int J Immunogenet 37: 367–372.
40. Kordonouri O, Hartmann R, Badenhoop K, Kahles H, Ilonen J (2010) PTPN22 1858T allele is associated with younger age at onset of type 1 diabetes and is not related to subsequent thyroid autoimmunity. Hum Immunol 71: 731–732.
41. Korolija M, Renar IP, Hadzija M, Medvidovic EP, Pavkovic P, et al. (2009) Association of PTPN22 C1858T and CTLA-4 A49G polymorphisms with Type 1 Diabetes in Croatians. Diabetes Res Clin Pract 86: e54–57.
42. Lavrikova E, Nikitin AG, Seregin Iu A, Zil'berman LI, Tsitlidze NM, et al. (2009) [Association of the C1858T polymorphism of the PTPN22 gene with type 1 diabetes]. Mol Biol (Mosk) 43: 1040–1043.
43. Maziarz M, Janer M, Roach JC, Hagopian W, Palmer JP, et al. (2010) The association between the PTPN22 1858C>T variant and type 1 diabetes depends on HLA risk and GAD65 autoantibodies. Genes Immun 11: 406–415.
44. Nielsen LB, Porksen S, Andersen ML, Fredheim S, Svensson J, et al. (2011) The PTPN22 C1858T gene variant is associated with proinsulin in new-onset type 1 diabetes. BMC Med Genet 12: 41.
45. Saccucci P, Del Duca E, Rapini N, Verrotti A, Piccinini S, et al. (2008) Association between PTPN22 C1858T and type 1 diabetes: a replication in continental Italy. Tissue Antigens 71: 234–237.
46. Smyth DJ, Cooper JD, Howson JM, Walker NM, Plagnol V, et al. (2008) PTPN22 Trp620 explains the association of chromosome 1p13 with type 1 diabetes and shows a statistical interaction with HLA class II genotypes. Diabetes 57: 1730–1737.
47. Steck AK, Baschal EE, Jasinski JM, Boehm BO, Bottini N, et al. (2009) rs2476601 T allele (R620W) defines high-risk PTPN22 type I diabetes-associated haplotypes with preliminary evidence for an additional protective haplotype. Genes Immun 10 Suppl 1: S21–26.

48. Taniyama M, Maruyama T, Tozaki T, Nakano Y, Ban Y (2010) Association of PTPN22 haplotypes with type 1 diabetes in the Japanese population. Hum Immunol 71: 795–798.
49. Zhebrun D, Kudryashova Y, Babenko A, Maslyansky A, Kunitskaya N, et al. (2011) Association of PTPN22 1858T/T genotype with type 1 diabetes, Graves' disease but not with rheumatoid arthritis in Russian population. Aging (Albany NY) 3: 368–373.
50. Nikitin AG, Lavrikova E, Seregin Iu A, Zil'berman LI, Tsitlidze NM, et al. (2010) [Association of the polymorphisms of the ERBB3 and SH2B3 genes with type 1 diabetes]. Mol Biol (Mosk) 44: 257–262.
51. Wang H, Jin Y, Reddy MV, Podolsky R, Liu S, et al. (2010) Genetically dependent ERBB3 expression modulates antigen presenting cell function and type 1 diabetes risk. PLoS One 5: e11789.
52. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, et al. (2007) A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 448: 591–594.
53. Wu X, Zhu X, Wang X, Ma J, Zhu S, et al. (2009) Intron polymorphism in the KIAA0350 gene is reproducibly associated with susceptibility to type 1 diabetes (T1D) in the Han Chinese population. Clin Endocrinol (Oxf) 71: 46–49.
54. Srinivasan S, Bolick DT, Lukashev D, Lappas C, Sitkovsky M, et al. (2008) Sphingosine-1-phosphate reduces CD4+ T-cell activation in type 1 diabetes through regulation of hypoxia-inducible factor short isoform I.1 and CD69. Diabetes 57: 484–493.
55. Thompson SD, Sudman M, Ramos PS, Marion MC, Ryan M, et al. (2010) The susceptibility loci juvenile idiopathic arthritis shares with other autoimmune diseases extend to PTPN2, COG6, and ANGPT1. Arthritis Rheum 62: 3265–3276.
56. Espino-Paisan L, De La Calle H, Fernandez-Arquero M, Figueredo MA, De La Concha EG, et al. (2011) Study of polymorphisms in 4q27, 10p15, and 22q13 regions in autoantibodies stratified type 1 diabetes patients. Autoimmunity 44: 624–630.
57. Zhernakova A, Alizadeh BZ, Bevova M, van Leeuwen MA, Coenen MJ, et al. (2007) Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. Am J Hum Genet 81: 1284–1288.
58. Lin G, Wang G, Liu G, Yang LJ, Chang LJ, et al. (2009) Treatment of type 1 diabetes with adipose tissue-derived stem cells expressing pancreatic duodenal homeobox 1. Stem Cells Dev 18: 1399–1406.
59. Noso S, Kataoka K, Kawabata Y, Babaya N, Hiromine Y, et al. (2010) Insulin transactivator MafA regulates intrathymic expression of insulin and affects susceptibility to type 1 diabetes. Diabetes 59: 2579–2587.
60. Douroudis K, Laine AP, Heinonen M, Hermann R, Lipponen K, et al. (2009) Association of CTLA4 but not ICOS polymorphisms with type 1 diabetes in two populations with different disease rates. Hum Immunol 70: 536–539.
61. Kawasaki E, Imagawa A, Makino H, Uga M, Abiru N, et al. (2008) Differences in the contribution of the CTLA4 gene to susceptibility to fulminant and type 1A diabetes in Japanese patients. Diabetes Care 31: 1608–1610.
62. Qu HQ, Bradfield JP, Grant SF, Hakonarson H, Polychronakos C (2009) Remapping the type I diabetes association of the CTLA4 locus. Genes Immun 10 Suppl 1: S27–32.
63. Saleh HM, Rohowsky N, Leski M (2008) The CTLA4 −819 C/T and +49 A/G dimorphisms are associated with Type 1 diabetes in Egyptian children. Indian J Hum Genet 14: 92–98.
64. Stechova K, Kolar M, Blatny R, Halbhuber Z, Vcelakova J, et al. (2011) Healthy first degree relatives of patients with type 1 diabetes exhibit significant differences in basal gene expression pattern of immunocompetent cells compared to controls: expression pattern as predeterminant of autoimmune diabetes. Scand J Immunol.
65. Kaas A, Pfleger C, Kharagjitsingh AV, Schloot NC, Hansen L, et al. (2011) Association between age, IL-10, IFNgamma, stimulated C-peptide and disease progression in children with newly diagnosed Type 1 diabetes. Diabet Med.
66. Liu Q, Sundar K, Mishra PK, Mousavi G, Liu Z, et al. (2009) Helminth infection can reduce insulitis and type 1 diabetes through CD25- and IL-10-independent mechanisms. Infect Immun 77: 5347–5358.
67. Petrich de Marquesini LG, Fu J, Connor KJ, Bishop AJ, McLintock NE, et al. (2010) IFN-gamma and IL-10 islet-antigen-specific T cell responses in autoantibody-negative first-degree relatives of patients with type 1 diabetes. Diabetologia 53: 1451–1460.
68. Sanda S, Roep BO, von Herrath M (2008) Islet antigen specific IL-10+ immune responses but not CD4+CD25+FoxP3+ cells at diagnosis predict glycemic control in type 1 diabetes. Clin Immunol 127: 138–143.
69. Settin A, Ismail A, El-Magd MA, El-Baz R, Kazamel A (2009) Gene polymorphisms of TNF-alpha-308 (G/A), IL-10(−1082) (G/A), IL-6(−174) (G/C) and IL-1Ra (VNTR) in Egyptian cases with type 1 diabetes mellitus. Autoimmunity 42: 50–55.
70. Zhu H, Qiu W, Lei P, Zhou W, Wen X, et al. (2008) IL-10 gene modified dendritic cells inhibit T helper type 1-mediated alloimmune responses and promote immunological tolerance in diabetes. Cell Mol Immunol 5: 41–46.
71. Alcina A, Fedetz M, Ndagire D, Fernandez O, Leyva L, et al. (2009) IL2RA/CD25 gene polymorphisms: uneven association with multiple sclerosis (MS) and type 1 diabetes (T1D). PLoS One 4: e4137.
72. Klinker MW, Schiller JJ, Magnuson VL, Wang T, Basken J, et al. (2010) Single-nucleotide polymorphisms in the IL2RA gene are associated with age at diagnosis in late-onset Finnish type 1 diabetes subjects. Immunogenetics 62: 101–107.
73. Maier LM, Lowe CE, Cooper J, Downes K, Anderson DE, et al. (2009) IL2RA genetic heterogeneity in multiple sclerosis and type 1 diabetes susceptibility and soluble interleukin-2 receptor production. PLoS Genet 5: e1000322.
74. Plagnol V (2010) [Haplotypes in the IL2RA region associated with diabetes and expression of the IL2RA protein on lymphocytes populations]. Med Sci (Paris) 26: 236–238.
75. Qu HQ, Bradfield JP, Belisle A, Grant SF, Hakonarson H, et al. (2009) The type I diabetes association of the IL2RA locus. Genes Immun 10 Suppl 1: S42–48.
76. Fichna M, Zurawek M, Fichna P, Januszkiewicz D, Nowak J (2011) Polymorphic variants of the IL2RA gene and susceptibility to type 1 diabetes in the Polish population. Tissue Antigens.
77. Steck AK, Liu SY, McFann K, Barriga KJ, Babu SR, et al. (2006) Association of the PTPN22/LYP gene with type 1 diabetes. Pediatr Diabetes 7: 274–278.
78. Kawasaki E, Awata T, Ikegami H, Kobayashi T, Maruyama T, et al. (2006) Systematic search for single nucleotide polymorphisms in a lymphoid tyrosine phosphatase gene (PTPN22): association between a promoter polymorphism and type 1 diabetes in Asian populations. Am J Med Genet A 140: 586–593.
79. Ladner MB, Bottini N, Valdes AM, Noble JA (2005) Association of the single nucleotide polymorphism C1858T of the PTPN22 gene with type 1 diabetes. Hum Immunol 66: 60–64.
80. Ide A, Kawasaki E, Abiru N, Sun F, Takahashi R, et al. (2002) Genetic association between interleukin-10 gene promoter region polymorphisms and type 1 diabetes age-at-onset. Hum Immunol 63: 690–695.
81. Metcalfe KA, Hitman GA, Pociot F, Bergholdt R, Tuomilehto-Wolf E, et al. (1996) An association between type 1 diabetes and the interleukin-1 receptor type 1 gene. The DiMe Study Group. Childhood Diabetes in Finland. Hum Immunol 51: 41–48.
82. Iwata I, Nagafuchi S, Nakashima H, Kondo S, Koga T, et al. (1999) Association of polymorphism in the NeuroD/BETA2 gene with type 1 diabetes in the Japanese. Diabetes 48: 416–419.
83. Ihara K, Ahmed S, Nakao F, Kinukawa N, Kuromaru R, et al. (2001) Association studies of CTLA-4, CD28, and ICOS gene polymorphisms with type 1 diabetes in the Japanese population. Immunogenetics 53: 447–454.
84. Anjos SM, Tessier MC, Polychronakos C (2004) Association of the cytotoxic T lymphocyte-associated antigen 4 gene with type 1 diabetes: evidence for independent effects of two polymorphisms on the same haplotype block. J Clin Endocrinol Metab 89: 6257–6265.
85. Ide A, Kawasaki E, Abiru N, Sun F, Kobayashi M, et al. (2004) Association between IL-18 gene promoter polymorphisms and CTLA-4 gene 49A/G polymorphism in Japanese patients with type 1 diabetes. J Autoimmun 22: 73–78.
86. Kamoun Abid H, Hmida S, Smaoui N, Kaabi H, Abid A, et al. (2001) [Association between type 1 diabetes and polymorphism of the CTLA-4 gene in a Tunisian population]. Pathol Biol (Paris) 49: 794–798.
87. Kikuoka N, Sugihara S, Yanagawa T, Ikezaki A, Kim HS, et al. (2001) Cytotoxic T lymphocyte antigen 4 gene polymorphism confers susceptibility to type 1 diabetes in Japanese children: analysis of association with HLA genotypes and autoantibodies. Clin Endocrinol (Oxf) 55: 597–603.
88. Lee YJ, Huang FY, Lo FS, Wang WC, Hsu CH, et al. (2000) Association of CTLA4 gene A-G polymorphism with type 1 diabetes in Chinese children. Clin Endocrinol (Oxf) 52: 153–157.
89. Mochizuki M, Amemiya S, Kobayashi K, Kobayashi K, Shimura Y, et al. (2003) Association of the CTLA-4 gene 49 A/G polymorphism with type 1 diabetes and autoimmune thyroid disease in Japanese children. Diabetes Care 26: 843–847.
90. Osei-Hyiaman D, Hou L, Zhiyin R, Zhiming Z, Yu H, et al. (2001) Association of a novel point mutation (C159G) of the CTLA4 gene with type 1 diabetes in West Africans but not in Chinese. Diabetes 50: 2169–2171.
91. Nishino M, Ikegami H, Fujisawa T, Kawaguchi Y, Kawabata Y, et al. (2005) Functional polymorphism in Z-DNA-forming motif of promoter of SLC11A1 gene and type 1 diabetes in Japanese subjects: association study and meta-analysis. Metabolism 54: 628–633.
92. Bassuny WM, Ihara K, Matsuura N, Ahmed S, Kohno H, et al. (2002) Association study of the NRAMP1 gene promoter polymorphism and early-onset type 1 diabetes. Immunogenetics 54: 282–285.
93. Nielsen C, Hansen D, Husby S, Jacobsen BB, Lillevang ST (2003) Association of a putative regulatory polymorphism in the PD-1 gene with susceptibility to type 1 diabetes. Tissue Antigens 62: 492–497.
94. Ongagna JC, Pinget M, Belcourt A (2005) Vitamin D-binding protein gene polymorphism association with IA-2 autoantibodies in type 1 diabetes. Clin Biochem 38: 415–419.
95. Gupta M, Nikitina-Zake L, Zarghami M, Landin-Olsson M, Kockum I, et al. (2003) Association between the transmembrane region polymorphism of MHC class I chain related gene-A and type 1 diabetes mellitus in Sweden. Hum Immunol 64: 553–561.
96. Ahmedov G, Ahmedova L, Sedlakova P, Cinek O (2006) Genetic association of type 1 diabetes in an Azerbaijanian population: the HLA-DQ, -DRB1*04, the insulin gene, and CTLA4. Pediatr Diabetes 7: 88–93.
97. Cinek O, Drevinek P, Sumnik Z, Bendlova B, Vavrinec J (2004) [Association of insulin gene variants with type 1 diabetes mellitus in Czech population]. Cas Lek Cesk 143: 318–322.

98. Noso S, Ikegami H, Fujisawa T, Kawabata Y, Asano K, et al. (2006) Association of SUMO4, as a candidate gene for IDDM5, with susceptibility to type 1 diabetes in Asian populations. Ann N Y Acad Sci 1079: 41–46.
99. Miyako K, Kohno H, Ihara K, Kuromaru R, Matsuura N, et al. (2004) Association study of human MTH1 gene polymorphisms with type 1 diabetes mellitus. Endocr J 51: 493–498.
100. Conway BR, Maxwell AP, Savage DA, Patterson CC, Doran PP, et al. (2004) Association between variation in the actin-binding gene caldesmon and diabetic nephropathy in type 1 diabetes. Diabetes 53: 1162–1165.
101. Vella A, Cooper JD, Lowe CE, Walker N, Nutland S, et al. (2005) Localization of a type 1 diabetes locus in the IL2RA/CD25 region by use of tag single-nucleotide polymorphisms. Am J Hum Genet 76: 773–779.
102. Guja C, Guja L, Nutland S, Rance H, Todd JA, et al. (2004) Strong association of insulin gene INS-VNTR polymorphisms with type 1 diabetes in the Romanian population. Rom J Intern Med 42: 313–323.
103. Perez De Nanclares G, Bilbao JR, Calvo B, Vitoria JC, Vazquez F, et al. (2003) 5′-Insulin gene VNTR polymorphism is specific for type 1 diabetes: no association with celiac or Addison's disease. Ann N Y Acad Sci 1005: 319–323.
104. Ide A, Kawasaki E, Abiru N, Sun F, Fukushima T, et al. (2003) Association of interleukin-18 gene promoter polymorphisms in type 1 diabetes and autoimmune thyroid disease. Ann N Y Acad Sci 1005: 436–439.
105. San-Pedro JI, Bilbao JR, Perez de Nanclares G, Vitoria JC, Martul P, et al. (2005) Heterogeneity of vitamin D receptor gene association with celiac disease and type 1 diabetes mellitus. Autoimmunity 38: 439–444.
106. Taverna MJ, Selam JL, Slama G (2005) Association between a protein polymorphism in the start codon of the vitamin D receptor gene and severe diabetic retinopathy in C-peptide-negative type 1 diabetes. J Clin Endocrinol Metab 90: 4803–4808.
107. Marti G, Audi L, Esteban C, Oyarzabal M, Chueca M, et al. (2004) [Association of vitamin D receptor gene polymorphism with type 1 diabetes mellitus in two Spanish populations]. Med Clin (Barc) 123: 286–290.
108. Tessier MC, Qu HQ, Frechette R, Bacot F, Grabs R, et al. (2006) Type 1 diabetes and the OAS gene cluster: association with splicing polymorphism or haplotype? J Med Genet 43: 129–132.
109. Lavery GG, McTernan CL, Bain SC, Chowdhury TA, Hewison M, et al. (2002) Association studies between the HSD11B2 gene (encoding human 11beta-hydroxysteroid dehydrogenase type 2), type 1 diabetes mellitus and diabetic nephropathy. Eur J Endocrinol 146: 553–558.
110. Nejentsev S, Guja C, McCormack R, Cooper J, Howson JM, et al. (2003) Association of intercellular adhesion molecule-1 gene with type 1 diabetes. Lancet 362: 1723–1724.
111. Nishimura M, Obayashi H, Maruya E, Ohta M, Tegoshi H, et al. (2000) Association between type 1 diabetes age-at-onset and intercellular adhesion molecule-1 (ICAM-1) gene polymorphism. Hum Immunol 61: 507–510.