Comparison between Hapgen and `waffect`

1 Protocol

For the reasons detailed in the fourth paragraph of the introduction, the retrospective strategies are the approaches most widely used to simulate genotypic data under H1. Among those, Hapgen (`http://www.stats.ox.ac.uk/~marchini/software/gwas/hapgen.html`) is a popular program renowned to have been used for power assessment in the comparison study of various available genotyping chips [7]. Hapgen is also reputed for its pioneering use in the evaluation and comparison of different methods devoted to the detection of genotype-phenotype associations at the level of the whole genome [3]. We compared Hapgen with `waffect`.

We compared `waffect` and Hapgen in the case of a single disease SNP and under various conditions. In particular, we considered the combination of three ingredients which were specified as input parameters for Hapgen and from which a corresponding input file (i.e. a vector of probabilities) was derived for `waffect`. The three ingredients are the following: the *minor allele frequency* (MAF) of the simulated disease SNP, the *prevalence disease*, and the severity of the disease expressed as *genotype relative risks* for various *genetic models*. The range of the MAF at the disease SNP was specified to be [0.2-0.3] or [0.1-0.2]. The prevalence was set to 0.01, a standard value used to simulate GWA data. Different genotype relative risks were considered as well as four genetic models: additive, dominant, recessive and multiplicative. Together with the genotype relative risks, the genetic models make it possible to specify the probability for a given subject to be affected, depending on her/his genotype at the susceptibility locus: $RR1 = \frac{prob(affected|Aa)}{prob(affected|aa)}$; $RR2 = \frac{prob(affected|AA)}{prob(affected|aa)}$, where A is the disease allele. For further details, Table 1 provides the correspondance between heterozygotous ($RR1$) and homozygotous ($RR2$) genotype relative risks under each genetic model.

genetic model	genotype relative risks			comment
	homozygous (aa)	heterozygous (Aa) RR1	homozygous (AA) RR2	
additive	1	$1+\frac{\alpha}{2}$	$1+\alpha$	additive effect
		1.3	1.6	0.3
		1.6	2.2	0.6
dominant	1	$1+\alpha$	$1+\alpha$	dominant effect
		1.3	1.3	0.3
		1.6	1.6	0.6
recessive	1	$1+\alpha$	$1+\alpha$	recessive effect
		1	1.3	0.3
		1	1.6	0.6
multiplicative	1	$1+\alpha$	$1+\alpha^2$	multiplicative effect
		1.3	1.69	1.3
		1.6	2.56	1.6

Table 1: Genotype relative risks under four disease models. A: disease allele. $RR1 = \frac{prob(affected|Aa)}{prob(affected|aa)} = \frac{f1}{f0}$; $RR2 = \frac{prob(affected|AA)}{prob(affected|aa)} = \frac{f2}{f0}$.

The Hapgen simulations were performed for 629 subjects (315 controls, 314 cases) and for 9,579 SNPs on chromosome 1. We chose to work with 9,579 SNPs for consistency and ease of comparison with the rest of our experimentations which were conducted on real data from the 1000 Genome Project. Indeed, after checking for MAF (at the 5% threshold) and Hardy-Weinberg equilibrium (at the 0.001 significance threshold), the number of SNPs from the 1000 Genome Project that we analyzed in each benchmark amounted roughly to 8,048. The 9,579 initial SNPs are located in the 13,372 Mb region delimited by loci 558,390 and 13930,000, in the reference Hapmap file (`http://hapmap.ncbi.nlm.nih.gov/`). Whatever the method, Hapgen

or `waffect`, the genotypic data were constructed based on the reference haplotypes of the HapMap phase II coming from U.S. residents of northern and western European ancestry (CEU). In the Hapgen specification, the disease SNP's MAF refers to the frequency in the reference panel. Finally, the second input required by Hapgen is the fine mapping of recombination rates across the chromosome region to be simulated.

Under each condition and for each of the two methods, Hapgen and `waffect`, $1,000$ simulations under $H1$ were generated together with $1,000$ simulations under $H0$. For each condition, the penetrances f_0, f_1 and f_2 specified in `waffect` for comparison purposes were obtained as follows: several Hapgen simulations under the appropriate parameter setting were performed, then the corresponding penetrances were averaged. Note that genotypes were only generated by Hapgen (under the H1 hypothesis). In the case of the Hapgen simulations, the H0 replicates were generated by permuting the phenotypes. Under each condition, the genotypic and phenotypic data generated by Hapgen were subject to a genome-wide association study. Similarly, together with the genotypes generated through Hapgen, the phenotypes affected through `waffect` under each condition were subject to a genome-wide association study. In the case of `waffect` simulations, we obtained the H0 replicates by affecting the phenotypes after specifying in `waffect` a uniform probability π_i accross all individuals. Again, each simulation under $H0$, performed through either Hapgen or `waffect` was subject to a genome-wide analysis.

Each of the $4,000$ datasets was analyzed in turn with the analysis toolset PLINK, a gold standard software dedicated to whole genome association studies (`http://pngu.mgh.harvard.edu/~purcell/plink/`) [4]. First, a routine quality control for genotypic data was carried out: SNPs with minor allele frequency less than 0.05 and SNPs deviant from the so-called Hardy-Weinberg equilibrium with a p-value below 0.001 were removed. We based our analysis on the Cochran-Armitage test for trend [1, 2]. Adapted after the chi-square test, this former test incorporates a suspected ordering in the effects of the categories of the genotype variable; it is widely used as a genotype-based test for case-control genetic association studies.

In contrast with simulations under H1, simulations under H0 are expected to provide few statistically significant results. This contrast can be observed through a binary (or two-class) classifier system. Typically, in a binary classification problem, an object has to be mapped into one of two classes, say 0 and 1. The analysis of the power to categorize objects with respect to two classes requires that the underlying reality is known. Generally, the parameters of a classifier algorithm are derived from training on known 0 and 1 examples; then the classifier is tested on (known) 0 and 1 examples that were not part of the training sets. Thus, four possible outcomes are identified: true positive (TP), true negative (TN), false positive (FP), false negative (FN), with, say class 0 considered as positive (P) and class 1 considered as negative (N). If an object is positive and classified as positive, it is counted as a TP; if it is classified as negative, it is counted as a FN. Comparing the true positive rate ($TPr = sensitivity = \frac{TP}{TP+FP} = TP/P$) with the false positive rate ($FPr = 1 - specificity = \frac{FP}{FP+TN} = FP/N$) is the key to the assessment of the classification power. In the technics of the Receiver Operating Characteristic (ROC) [6], a curve is generated by plotting $sensitivity$ on the y axis as a function of $1 - specificity$ on the x axis for a continuum of diagnostic criteria.

Shifting from 1 and 0 denominations to $H0$ and $H1$ labels, and relying on a relevant diagnostic criterion, our aim is to compare the ROC curves plotted for $2,000$ H1 and H0 Hapgen simulations and for $2,000$ H1 and H0 `waffect` simulations respectively, under the same conditions (MAF, RR1 - RR2/genetic model). In our study, the comparison was performed twice, based on the two following statistics - or diagnostic criteria - :

- $S1 = \max_{j \in \mathscr{J}_\rho} -(\log_{10} p_j)$,
- $S2 = \mid j \in \mathscr{J}_\rho, p_j \leq \alpha_0 \mid$,

where p_j denotes the p-value of SNP j and $\mathscr{J}_\rho$ denotes the subset of SNPs such that the distance between their loci and the locus of a disease SNP is less than a radius ρ (hereafter called window half-width). The window is centered on a reference locus, namely the disease locus under H1. The threshold α_0 was set to 0.1. For each statistic, the continuum of statistic values was obtained over the $2,000$ PLINK runs and encompassed at most $2,000$ values, from minimum to maximum. In order to further derive false positive rates and true positive

rates, each value of this continuum was successively set as the threshold used to discriminate between H0 and H1.

For each genetic model, we considered two MAF intervals, combined with two risk levels. We considered risk levels reflecting four genetic diseases (additive, dominant, recessive, multiplicative). Therefore, for the sake of the comparison of `waffect` and Hapgen, we generated no less than $64,000$ simulations ($4 \times 2 \times 2 \times 4,000$). We ran as many association studies with the PLINK package.

For each of the two statistics, we plotted eight ROC curves, setting the half-width ρ of the window successively to 0 kb, 1 kb, 5 kb, 10 kb, 20 kb, 50 kb, 100 kb and "*inf*". In the latter case, for each PLINK execution, the statistic is calculated considering the largest (*i.e.* the whole) window. Fixing the simulation conditions (MAF, genetic model, RR1 - RR2), the statistic and the window half-width parameter ρ, we compared the ROC curves obtained from the two respective sets of $2,000$ Hapgen and $2,000$ `waffect` simulations. In complement to the graphical trends, we also compared the areas under the curves (AUCs) obtained with Hapgen and `waffect` respectively. We used one of the most recent R packages dedicated to ROC curve plotting and comparison [5] (`http://cran.r-project.org/web/packages/pROC/index.html`).

2 Hapgen simulation versus `waffect` simulation

Hapgen requires a reference panel of haplotypes together with a fine map of recombination rates whereas `waffect` only needs genotypic data. Furthermore, when replicating Hapgen simulations under the same specified conditions, there is no guarantee that the conditions are strictly kept constant throughout the whole simulation.

First, it has to be noted that even if the number of SNPs specified to Hapgen was constant across the replicates, the effective number of SNPs analyzed through PLINK varied through these replicates. Figure 1 illustrates this fact. The explanation for this variation relies on the change in the disease SNP's locus accross Hapgen simulations. The genetic data simulated by Hapgen are generated as follows: for a given individual, once the alleles at the disease locus have been fixed, a pair of haplotypes is generated, conditionally to the former alleles and to the genetic recombination map known for the chromosome region concerned. Starting from, say, r haplotypes already simulated, the hidden state X_j of an HMM model specifies which allele is selected for the locus j of the $r+1^{th}$ haplotype, amongst the alleles of the r haplotypes. In particular, the transitions between the states X_{j+1} and X_j of the Markov chain are ruled by the physical distance and the mean recombination rate between loci j and $j+1$. Then Hapgen implements two similar processes which extend the haplotype under construction to the right and left respectively. Moreover, partial copies of haplotype subregions are blurred through the simulation of mutations. Even if two simulation processes started from the same disease locus, these two stochastic aspects (transitions through Xjs and mutations) would generate two different benchmarks. Therefore, as two Hapgen runs select randomly the susceptibility loci (provided that they satisfy the MAF constraint), in practice, it is highly unlikely that each SNP would have the same genotype frequencies in both the two benchmarks being generated. Now the data preprocessing, that checks for MAF and Hardy-Weinberg equilibrium, depends on these genotype frequencies. Thus, the number of excluded SNPs and, symmetrically the number of SNPs analyzed through PLINK cannot be constant for all simulations.

Figure 2 describes the genotype frequencies across $1,000$ Hapgen simulations.

On the other hand, not only did the number of SNPs vary through Hapgen replicates, but the disease SNP's MAF also fluctuated across replicates. The explanation is that it is impossible to specify to Hapgen a MAF single value for the disease SNP. Instead, a MAF interval has to be provided, so that there is a chance to encounter a SNP checking the MAF constraint in the Hapmap reference panel. Note that throughout all our simulations, the MAF constraint was always satisfied, which prevented Hapgen to carry out a full random choice. In addition, since Hapgen simulation proceeds stochastically, the disease SNP's MAF in the reference panel and the effective disease SNP's MAF observed in the simulated data may differ, the latter possibly ending up outside the specified range (see Figure 3).

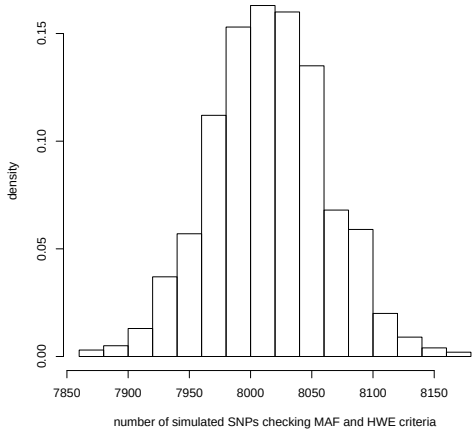


Figure 1: Histogram describing the distribution of the number of SNPs generated through Hapgen and checking MAF and Hardy-Weinberg equilibrium criteria, across $1,000$ simulations, and after specifying an initial number of SNPs equal to $9,579$. $MAF \in [0.2-0.3]$, $RR1 = 1.6$, $RR2 = 2.2$.

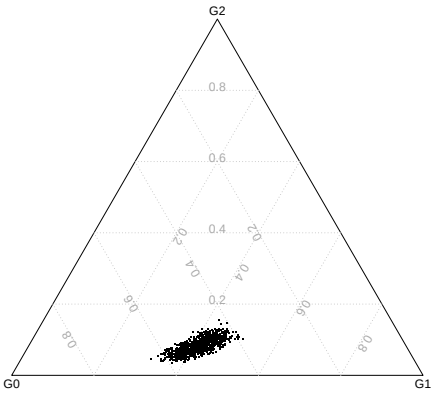


Figure 2: Ternary plot describing the distribution of the frequencies of genotypes for the SNPs generated through Hapgen and checking MAF and Hardy-Weinberg equilibrium criteria, across $1,000$ simulations. $MAF \in [0.2-0.3]$, $RR1 = 1.6$, $RR2 = 2.2$. *Gi* denotes the genotype with *i* rare alleles.

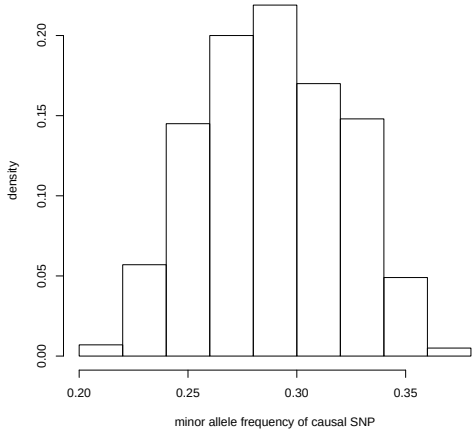
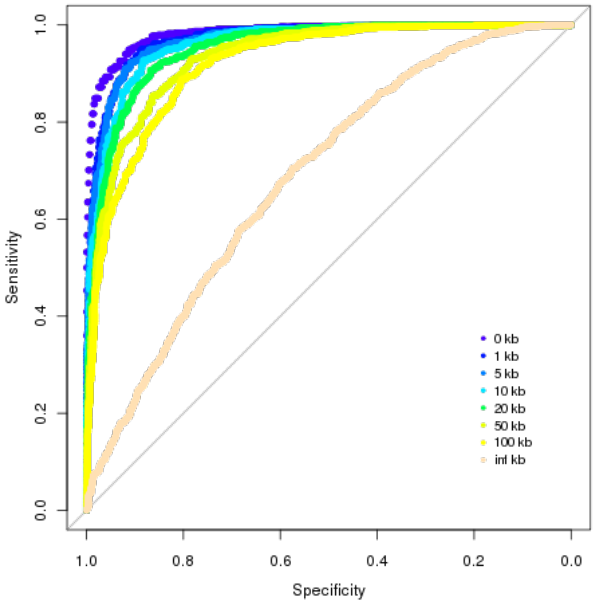


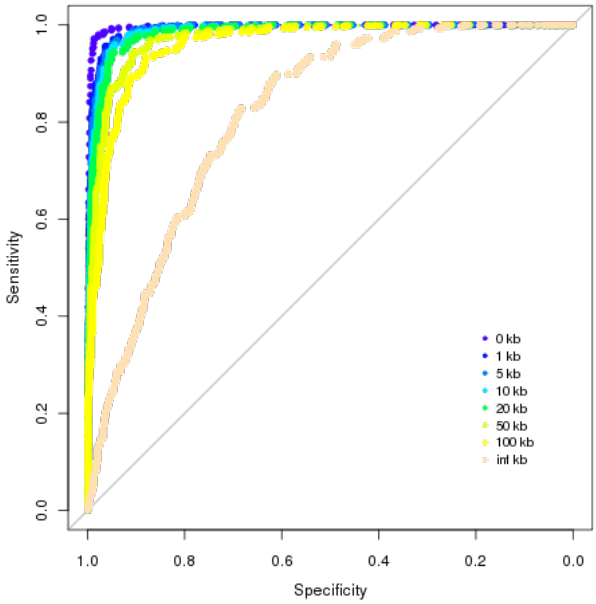
Figure 3: Histogram describing the distribution of the effective MAF of the susceptibility SNP, across $1,000$ simulations, for a MAF specified in $[0.2-0.3]$.

3 Additive genetic model

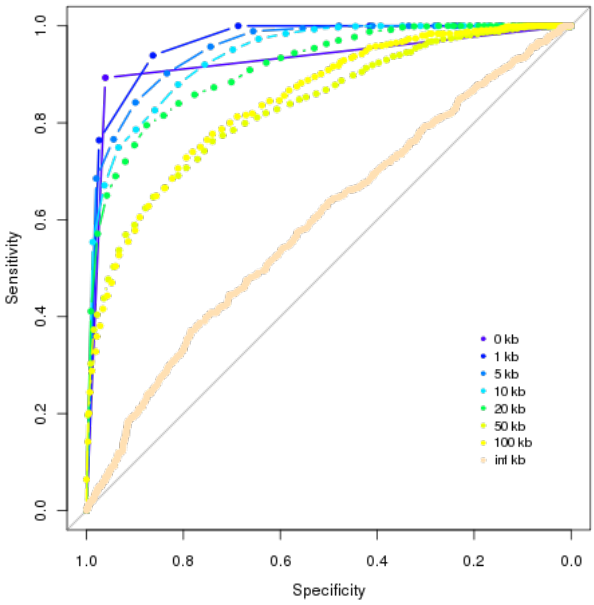
3.1 Simulations for MAF in $[0.2-0.3]$, additive model, $RR1 = 1.6$, $RR2 = 2.2$



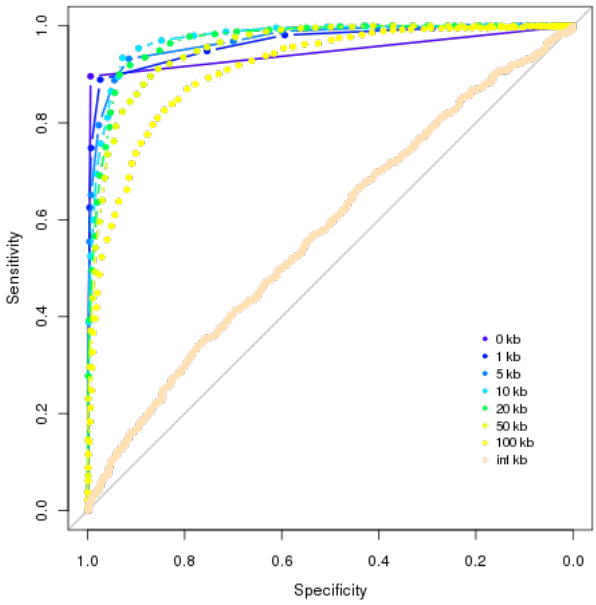
(a) Hapgen - statistic S1



(b) `waffect`- statistic S1



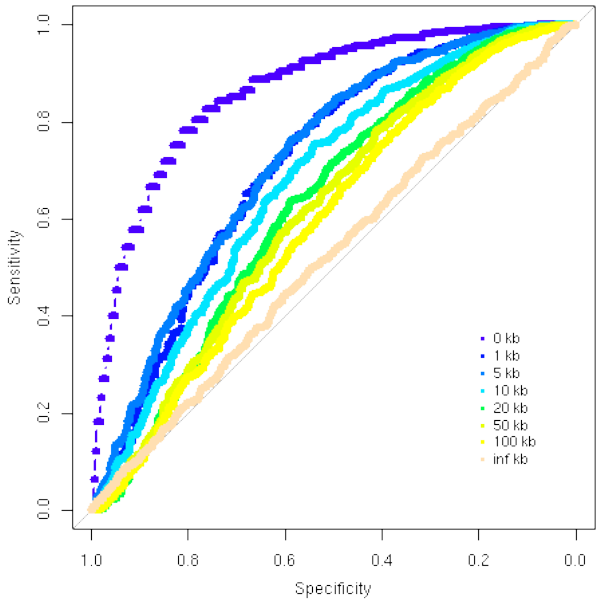
(c) Hapgen - statistic S2



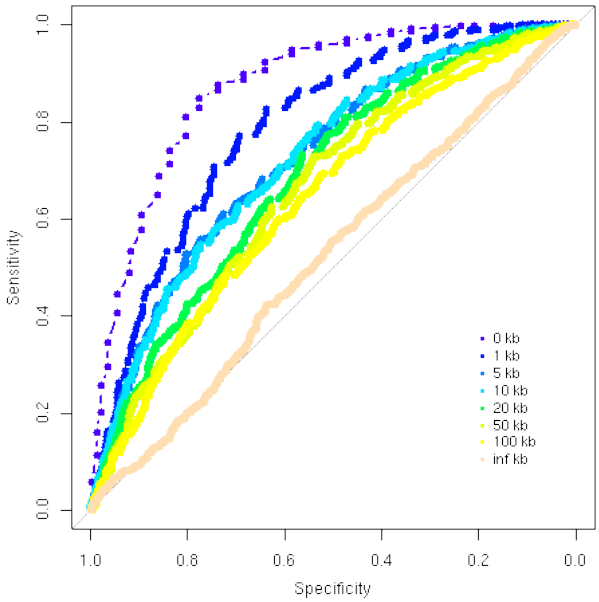
(d) `waffect`- statistic S2

Figure 4: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for the two statistics S1 and S2 - $MAF \in [0.2-0.3]$, additive model, $RR1 = 1.6$, $RR2 = 2.2$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

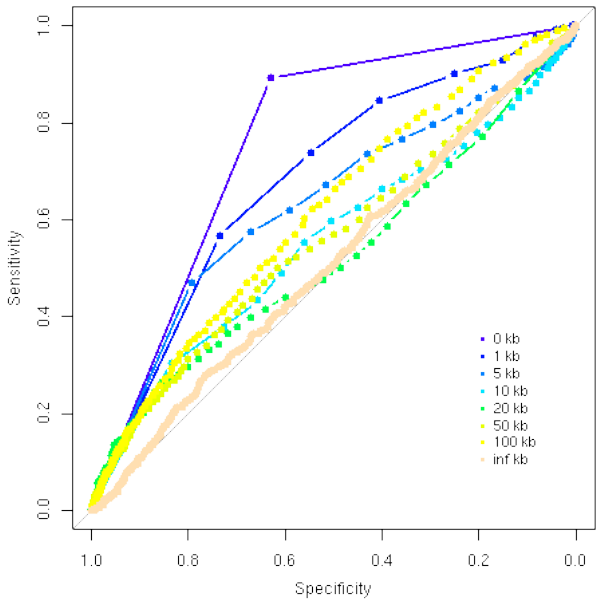
3.2 Simulations for MAF in $[0.2-0.3]$, additive model, $RR1 = 1.3$, $RR2 = 1.6$



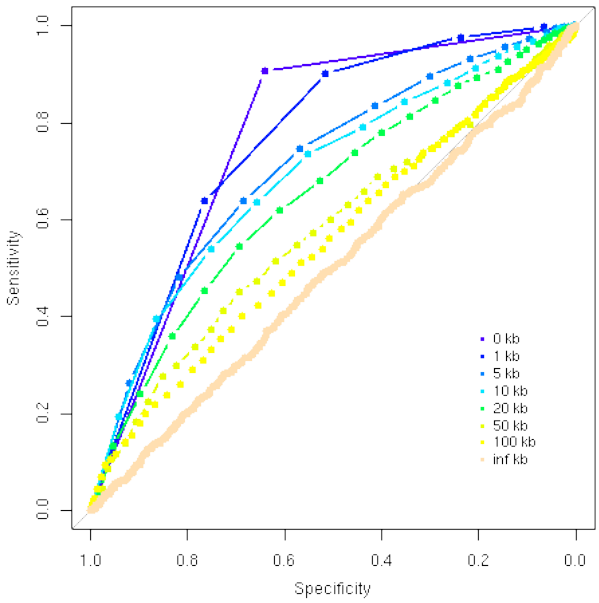
(a) Hapgen - statistic S1



(b) `waffect`- statistic S1



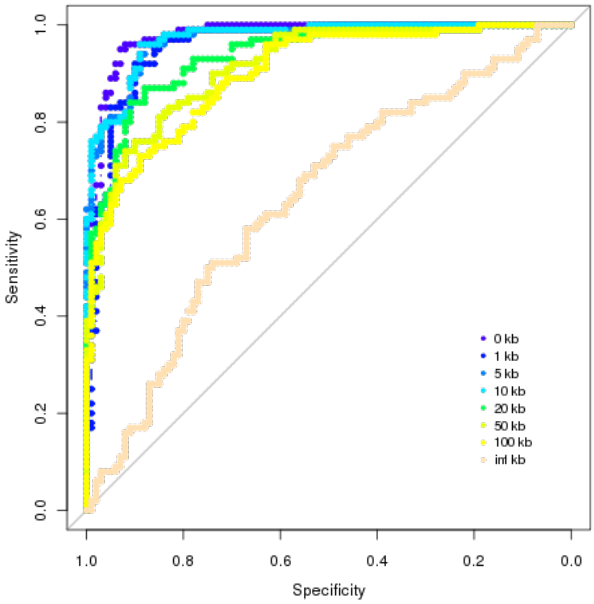
(c) Hapgen - statistic S2



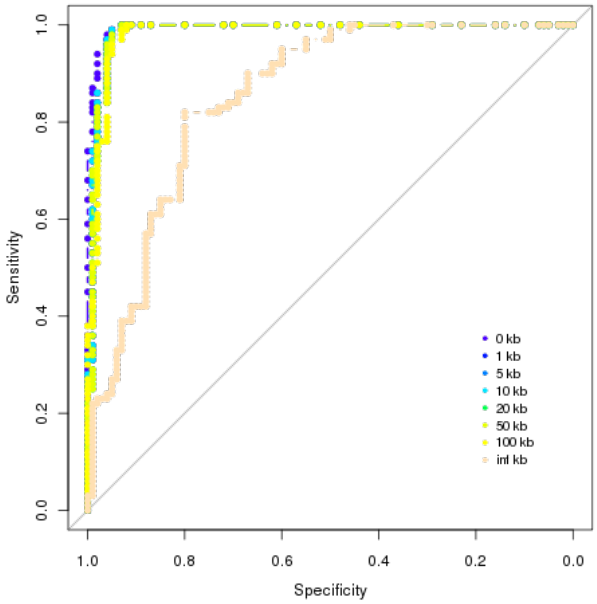
(d) `waffect`- statistic S2

Figure 5: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for the two statistics S1 and S2 - $MAF \in [0.2-0.3]$, additive model, $RR1 = 1.3$, $RR2 = 1.6$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

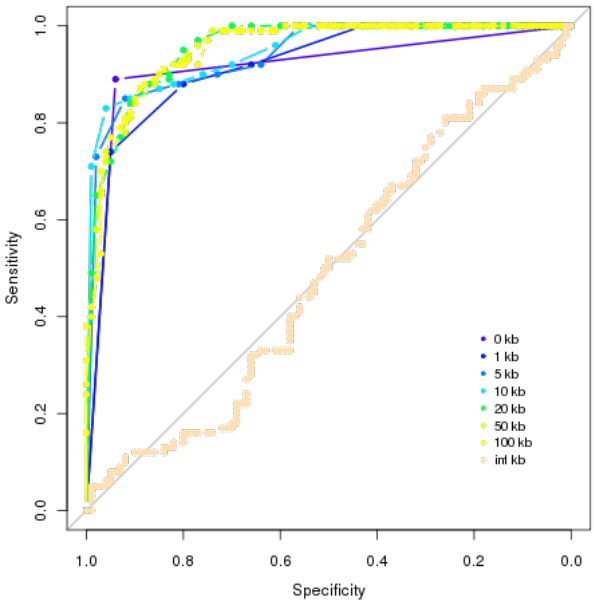
3.3 Simulations for MAF in $[0.1-0.2]$, additive model, $RR1 = 1.6$, $RR2 = 2.2$



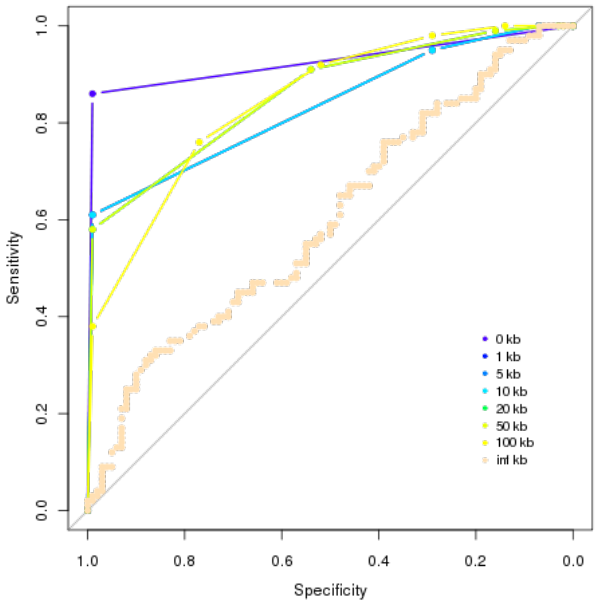
(a) Hapgen - statistic S1



(b) `waffect`- statistic S1



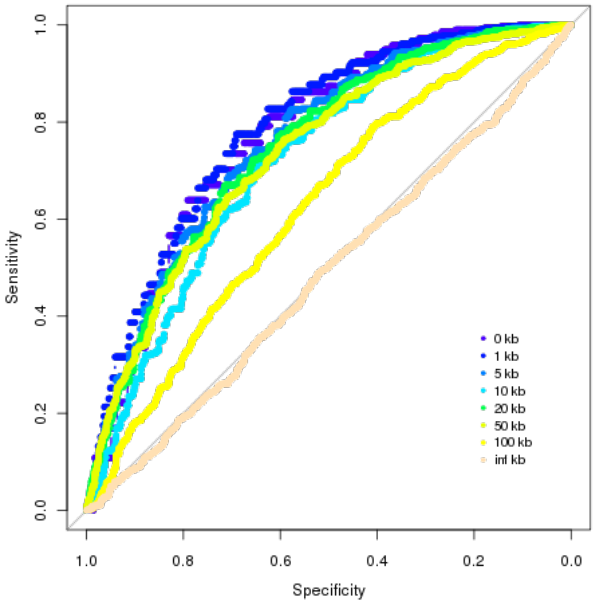
(c) Hapgen - statistic S2



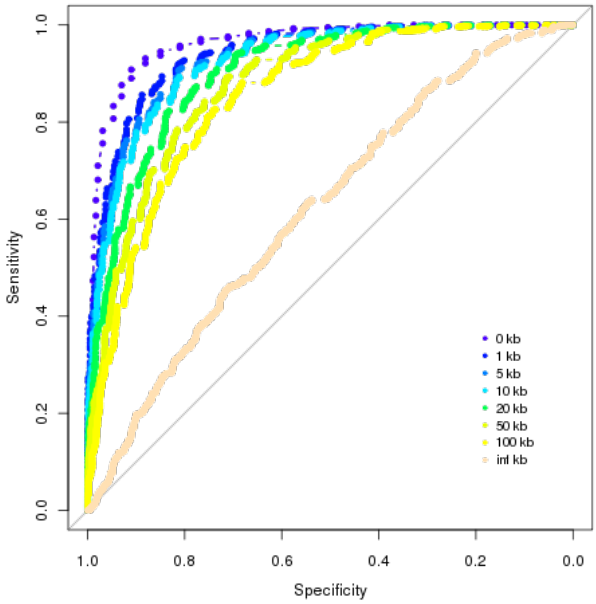
(d) `waffect`- statistic S2

Figure 6: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.1-0.2]$, additive model, $RR1 = 1.6$, $RR2 = 2.2$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

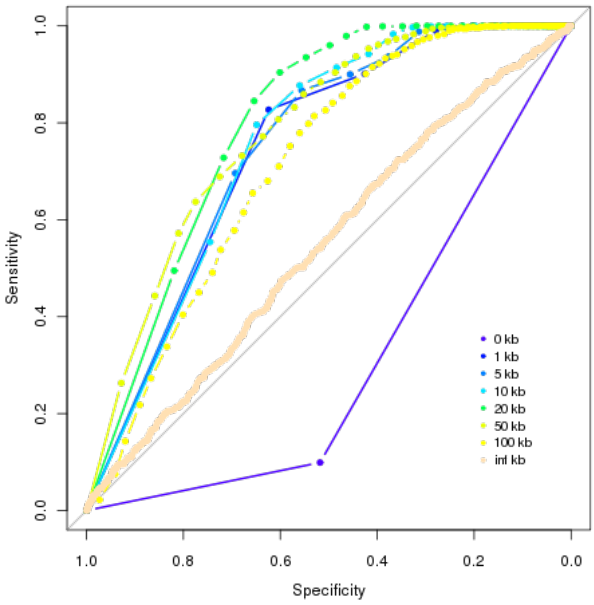
3.4 Simulations for MAF in $[0.1-0.2]$, additive model, $RR1 = 1.3$, $RR2 = 1.6$



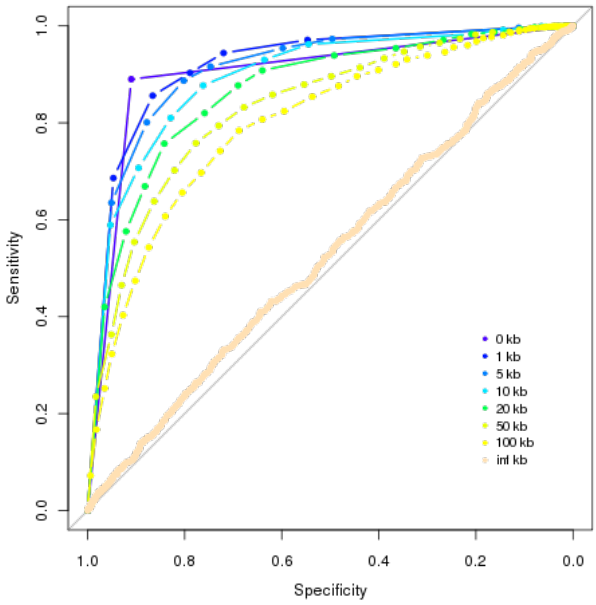
(a) Hapgen - statistic S1



(b) `waffect`- statistic S1



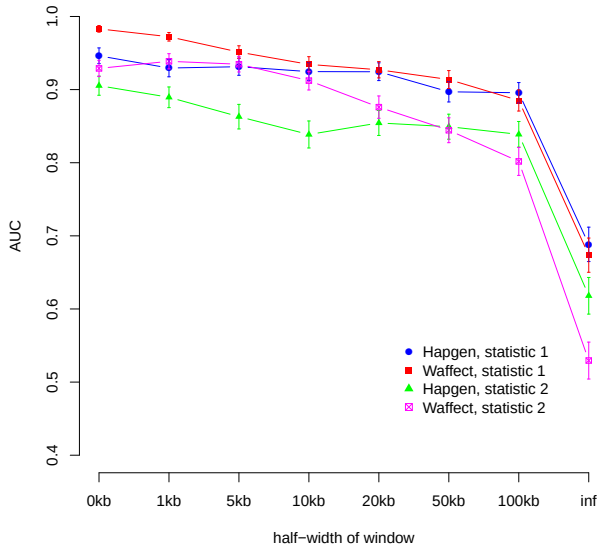
(c) Hapgen - statistic S2



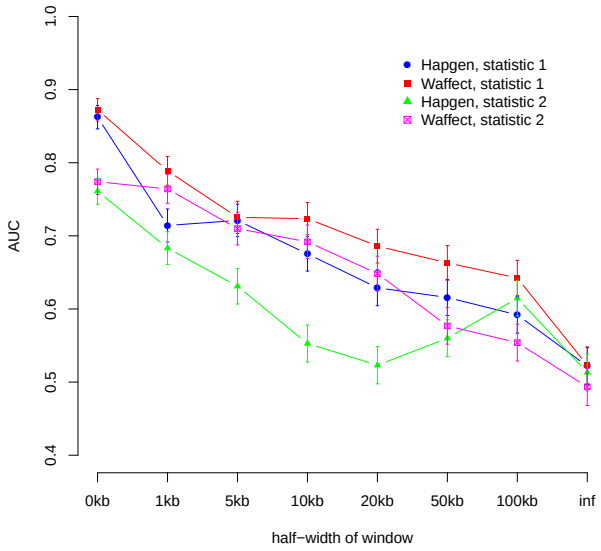
(d) `waffect`- statistic S2

Figure 7: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.1-0.2]$, additive model, $RR1 = 1.3$, $RR2 = 1.6$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

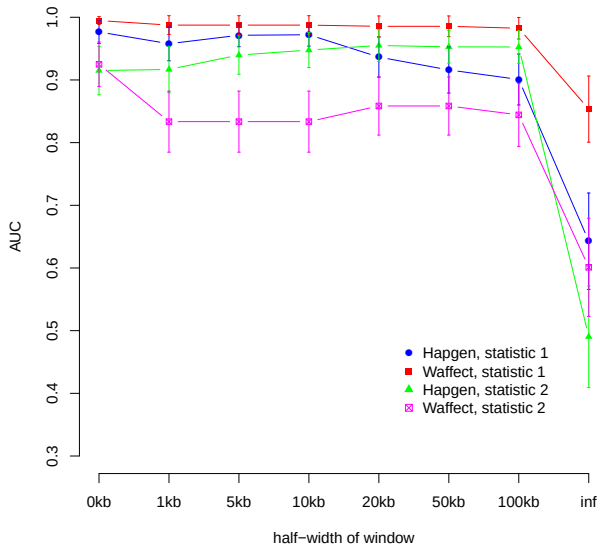
3.5 Recapitulation



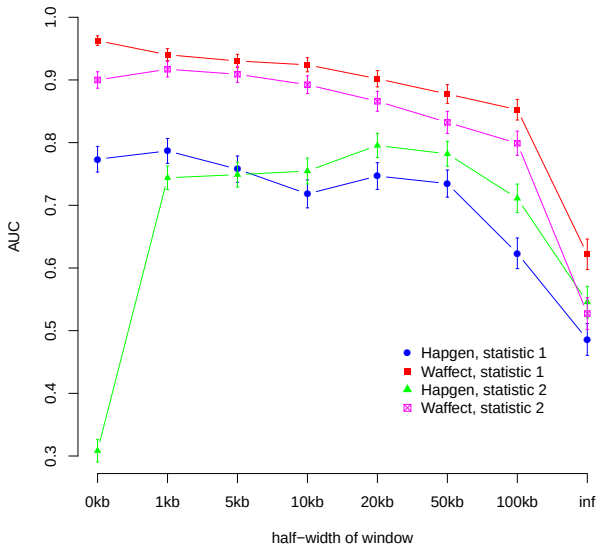
(a) $MAF \in [0.2-0.3]$; $RR1 = 1.6$; $RR2 = 2.2$



(b) $MAF \in [0.2-0.3]$; $RR1 = 1.3$; $RR2 = 1.6$



a) $MAF \in [0.1-0.2]$; $RR1 = 1.6$; $RR2 = 2.2$



(b) $MAF \in [0.1-0.2]$; $RR1 = 1.3$; $RR2 = 1.6$

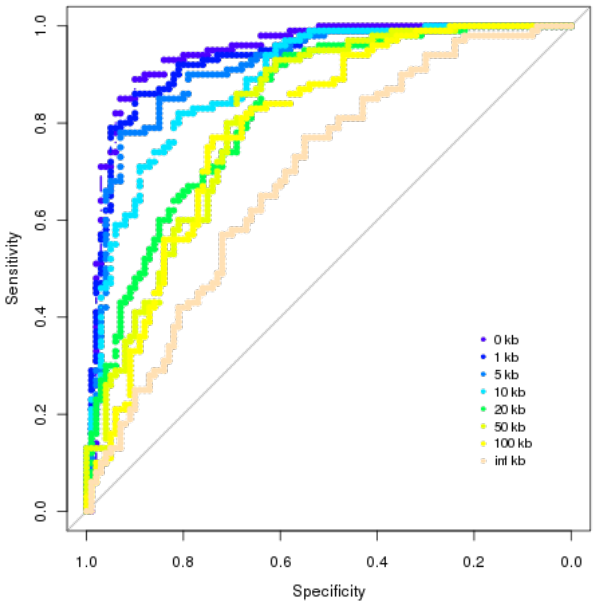
Figure 8: Comparison of the area under the curves (AUCs) corresponding to the ROC curves of figures 4, 5, 6 and 7.

Discussion Regarding both statistics, in all four simulated scenarios, we can observe that the trends for the discrimination power between H0 and H1 are similar for `waffect` and Hapgen. In this additive case, a more thorough analysis shows that the discrimination power with `waffect` is greater than that with Hapgen. A notable exception is encountered for the case of low MAF and high risk ($RR1 = 1.6$). Nevertheless, in all other three cases, an inversion of the trend can be observed beyond 100 kb or 50 kb (high MAF, low risk). Besides, a constant trend is observable: the discrepancy between the Hapgen and `waffect` curves is more important for statistic S2 than for statistic S1 (with the exception of the low MAF and weak risk). Regarding the high risk, and whatever the MAF, the common trend for both methods is a high discrimination power; only beyond 100 kb does the discrimination power drastically drop. In contrast, for a low risk, the trend is a slope from 0.7-0.9

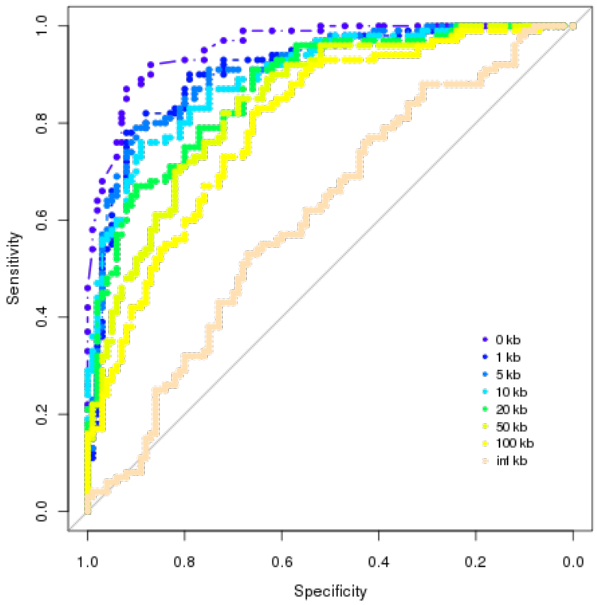
(high MAF) (resp. 0.8-0.95 (low MAF)) to 0.5. The lower discrimination power was expected for the weak risk with respect to a highest risk.

4 Dominant genetic model

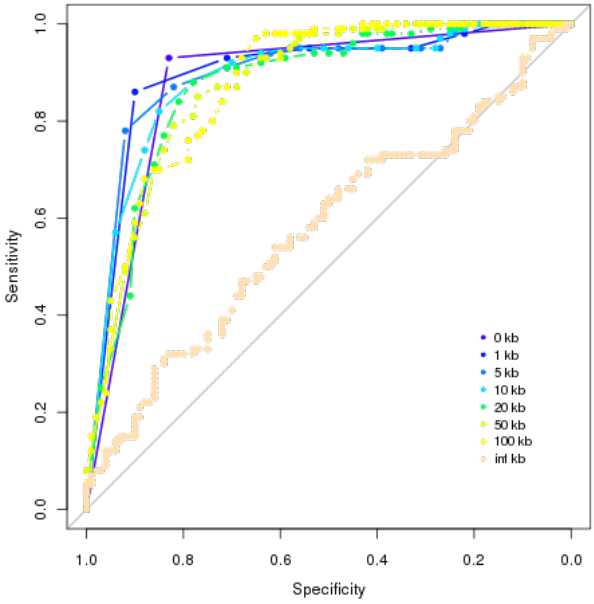
4.1 Simulations for MAF in $[0.2-0.3]$, dominant model, $RR1 = RR2 = 1.6$



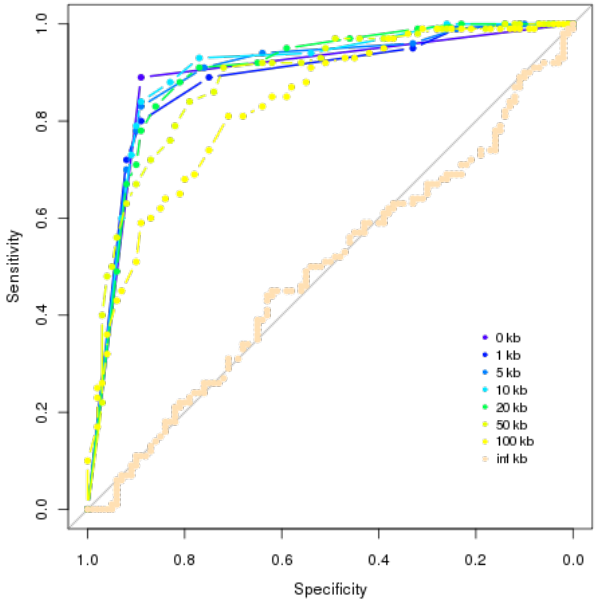
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



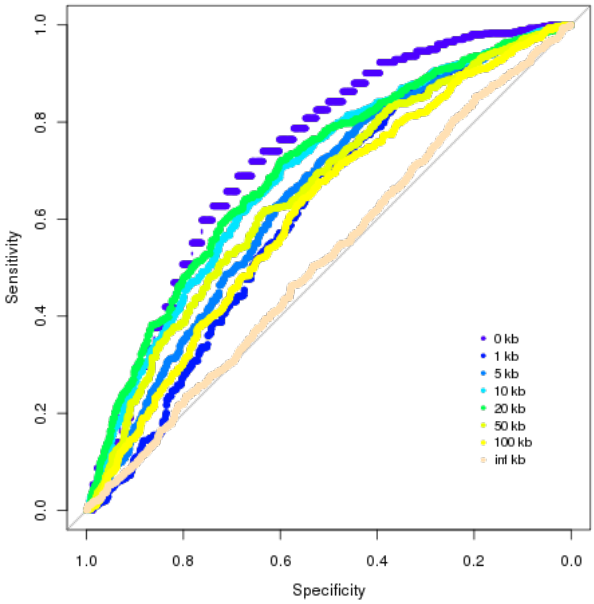
(c) Hapgen - statistic S2



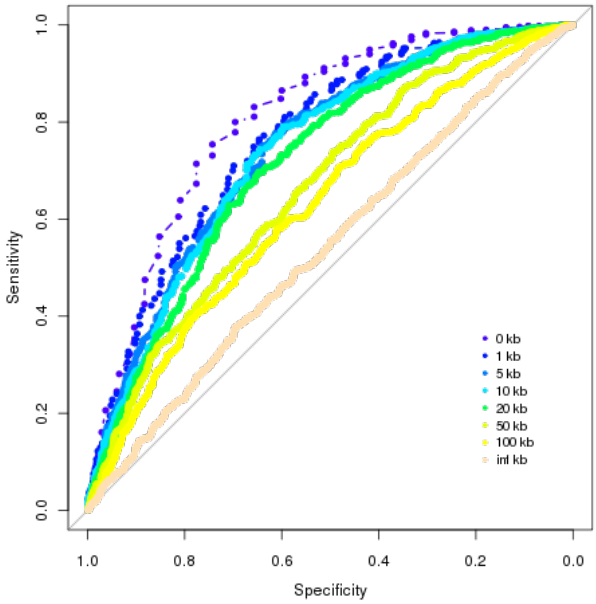
(d) Waffect - statistic S2

Figure 9: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for the two statistics S1 and S2 - $MAF \in [0.2-0.3]$, dominant model, $RR1 = RR2 = 1.6$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

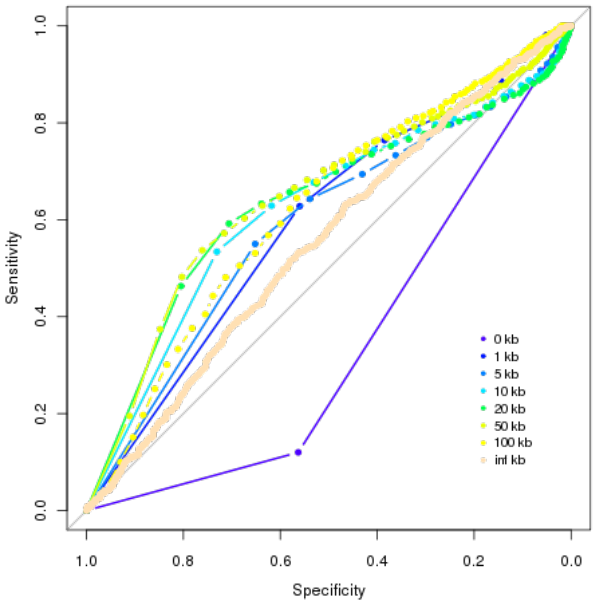
4.2 MAF in $[0.2-0.3]$, dominant model, $RR1 = RR2 = 1.3$



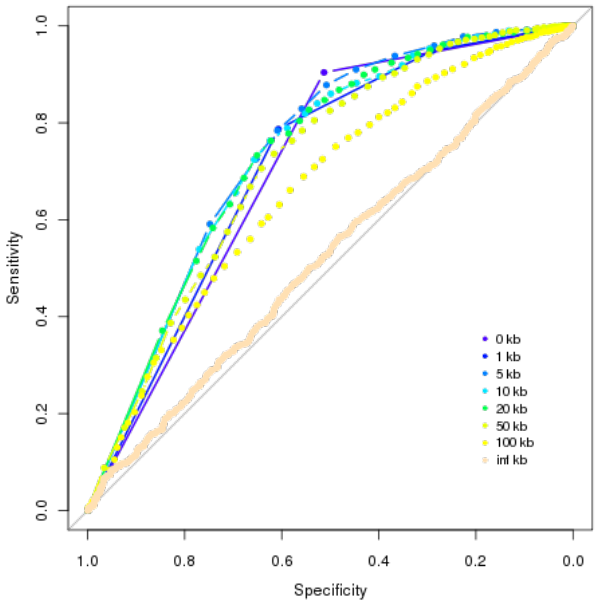
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



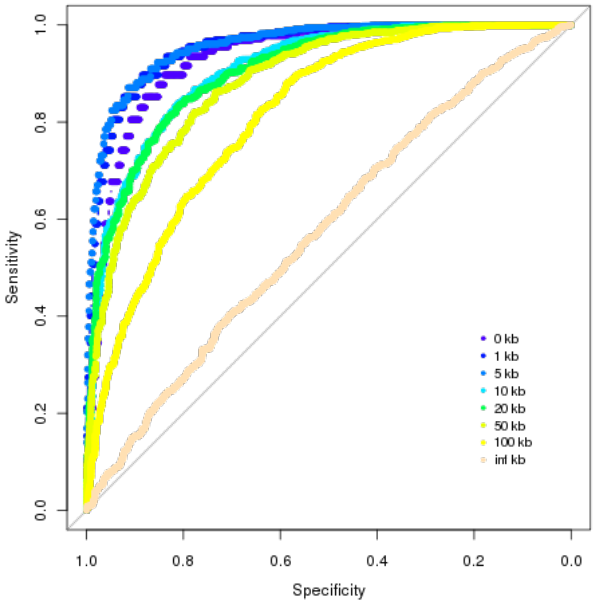
(c) Hapgen - statistic S2



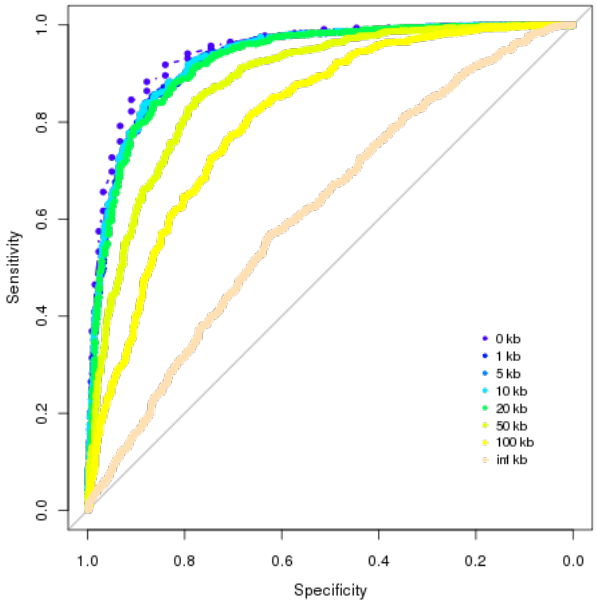
(d) Waffect - statistic S2

Figure 10: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.2-0.3]$, dominant model, $RR1 = RR2 = 1.3$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

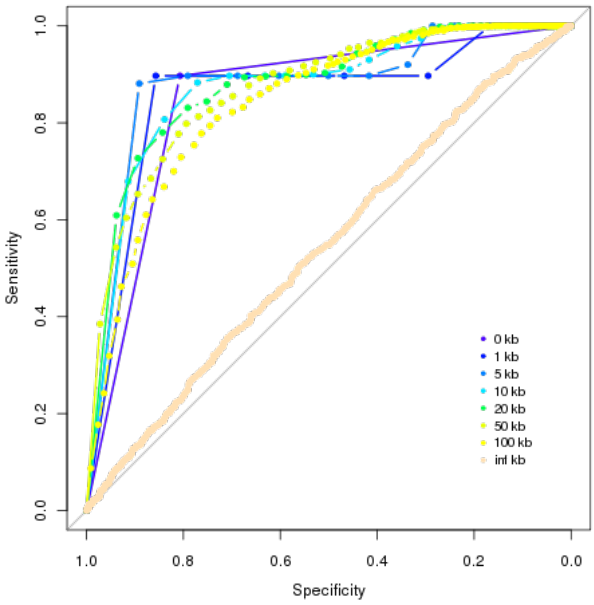
4.3 MAF in $[0.1 - 0.2]$, dominant model, $RR1 = RR2 = 1.6$



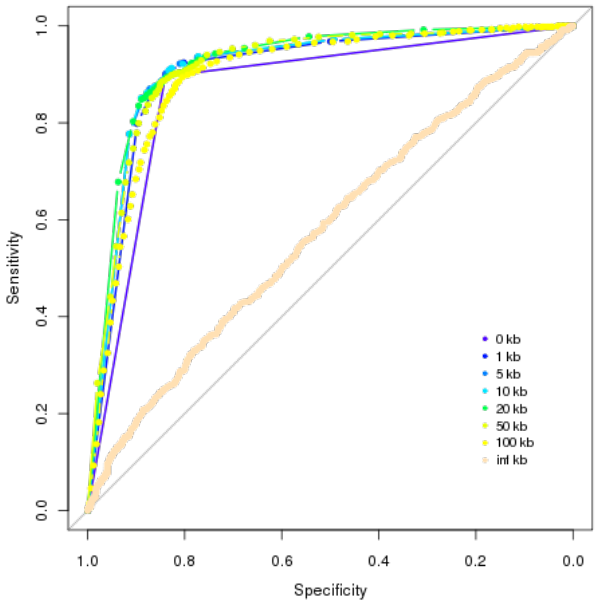
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



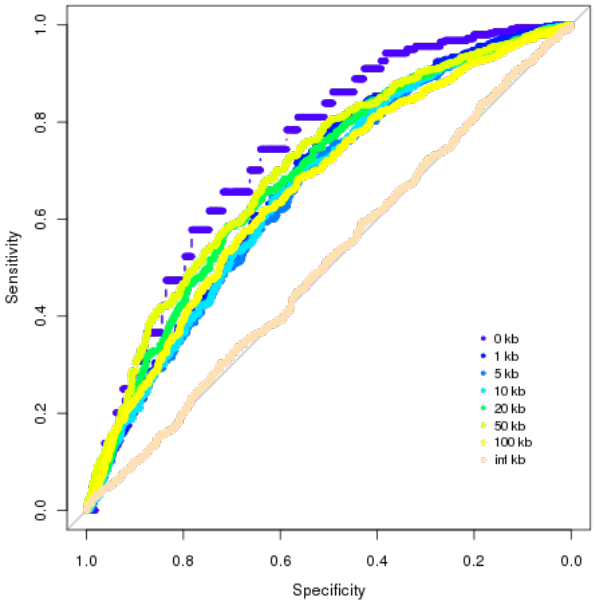
(c) Hapgen - statistic S2



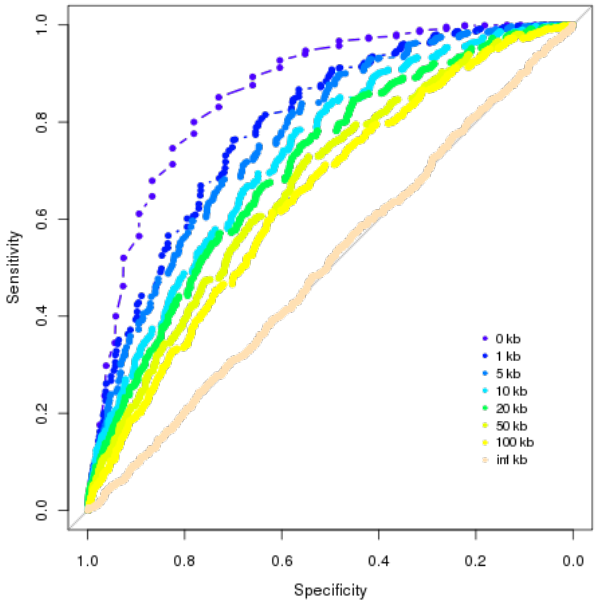
(d) Waffect - statistic S2

Figure 11: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.1 - 0.2]$, dominant model, $RR1 = RR2 = 1.6$. S1 is defined as the maximum $-log_{10}(p - value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

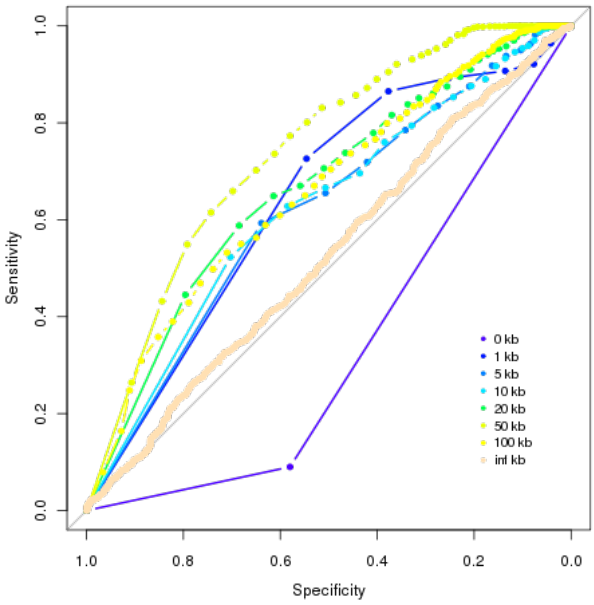
4.4 MAF in $[0.1-0.2]$, dominant model, $RR1 = RR2 = 1.3$



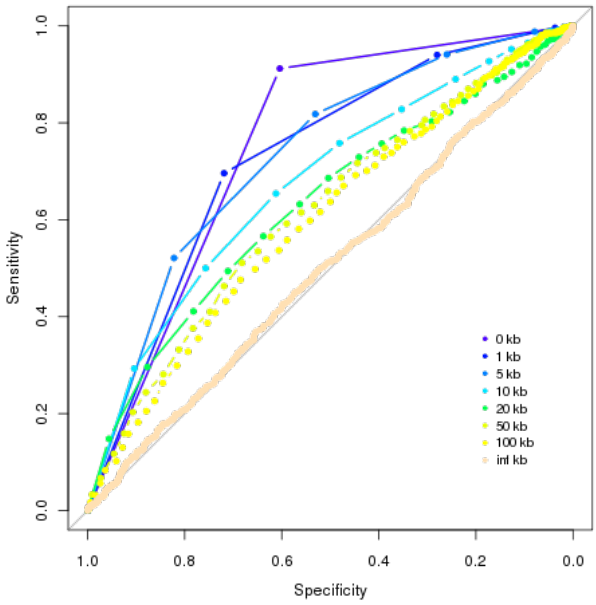
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



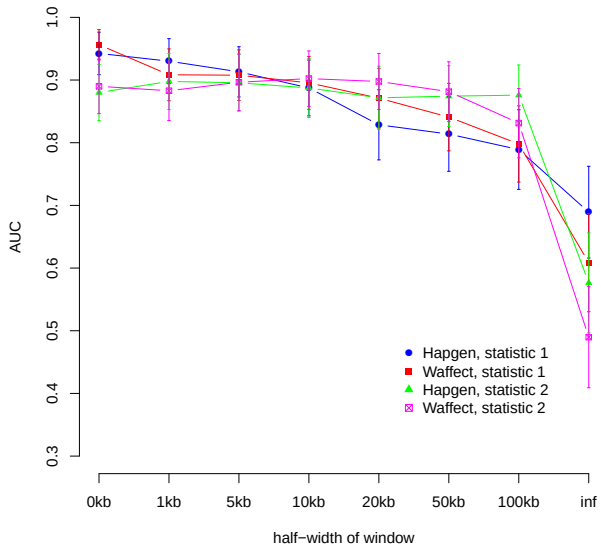
(c) Hapgen - statistic S1



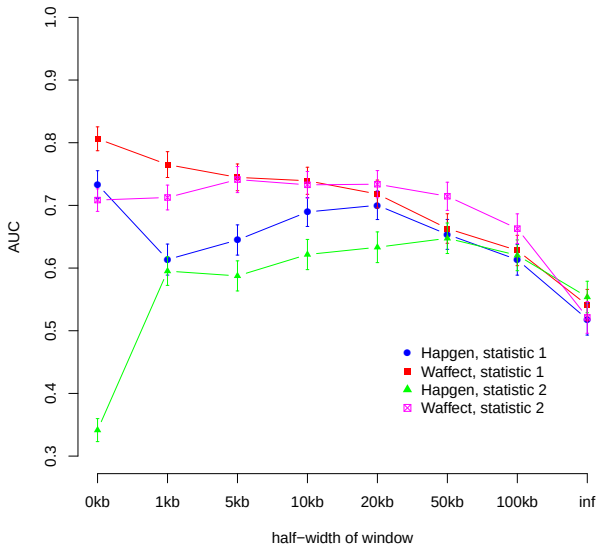
(d) Waffect - statistic S1

Figure 12: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.1-0.2]$, dominant model, $RR1 = RR2 = 1.3$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

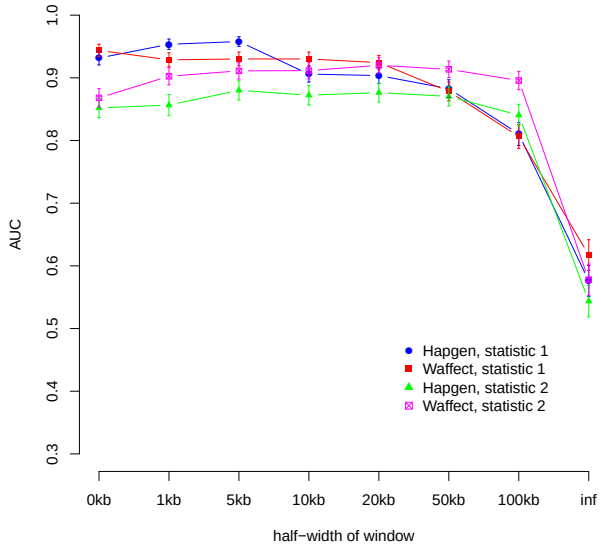
4.5 Recapitulation



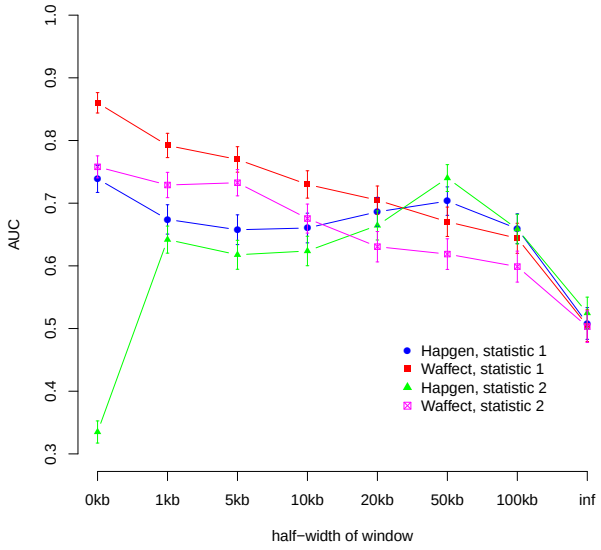
(a) $MAF \in [0.2-0.3]$; $RR1 = RR2 = 1.6$



(b) $MAF \in [0.2-0.3]$; $RR1 = RR2 = 1.3$



a) $MAF \in [0.1-0.2]$; $RR1 = RR2 = 1.6$



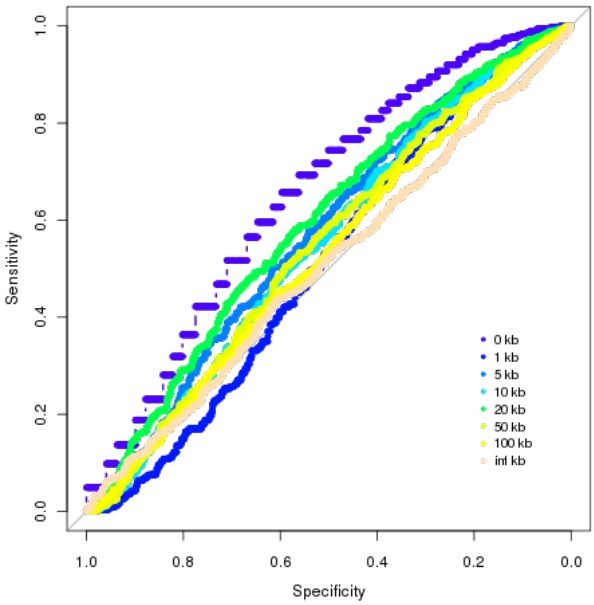
(b) $MAF \in [0.1-0.2]$; $RR1 = RR2 = 1.3$

Figure 13: Comparison of the area under the curves (AUCs) corresponding to the ROC curves of figures 9, 10, 11 and 12.

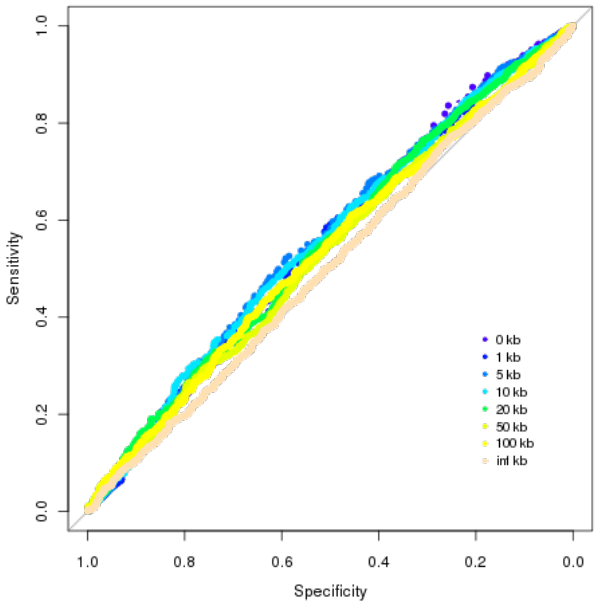
Discussion As in the additive case, trends remain similar for `waffect` and Hapgen: on the one hand, the discrimination power decreases as the risk decreases; on the other hand, for a fixed risk, the power decreases as the window width increases. However, in the dominant case, the comparison calls for some specific comments. First, it has to be noted that in five out of the eight cases considered (2 statistics $\times$ 4 conditions), neither method constantly overcomes the other one: whatever the statistic considered, the Hapgen and Waffect curves either overlap (high MAF, high risk) or show an intersection beyond which one of the method overcomes the other in terms of discrimination. In contrast, in the three remaining cases, the comparison is always in favour of `waffect` (statistic S1, high MAF, low risk; statistic S2, high MAF and low risk and *vice versa*).

5 Recessive genetic model

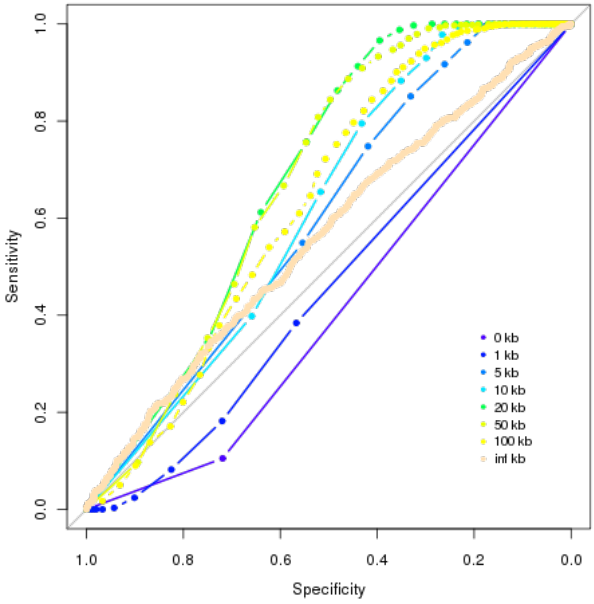
5.1 Simulations for MAF in $[0.2-0.3]$, recessive model, $RR1 = 1.0$, $RR2 = 1.6$



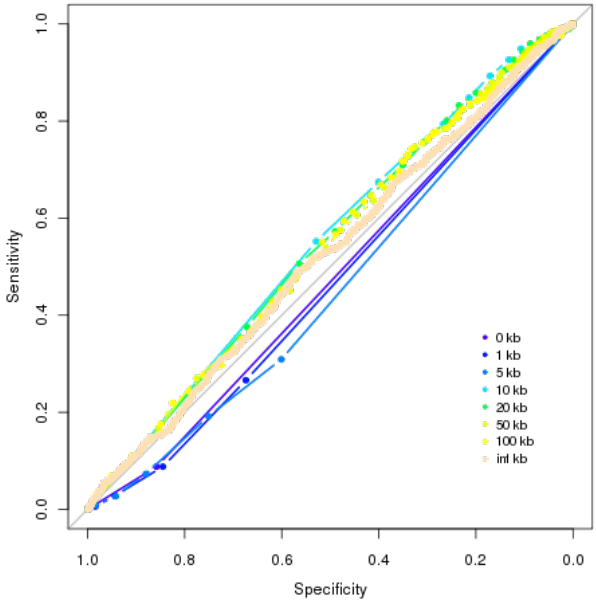
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



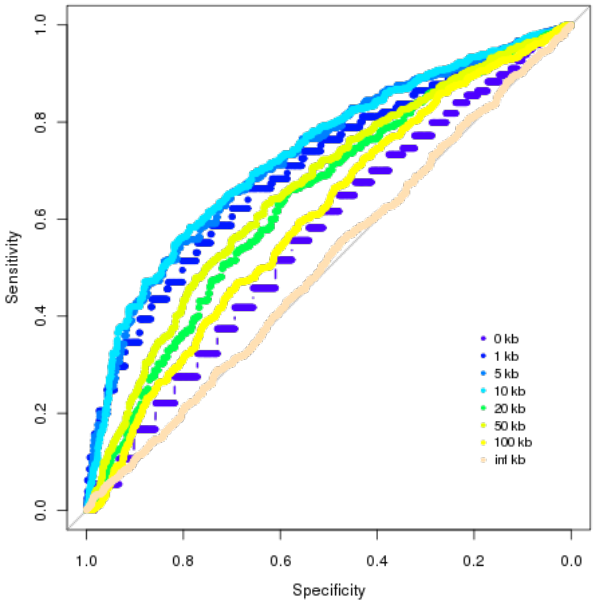
(c) Hapgen - statistic S2



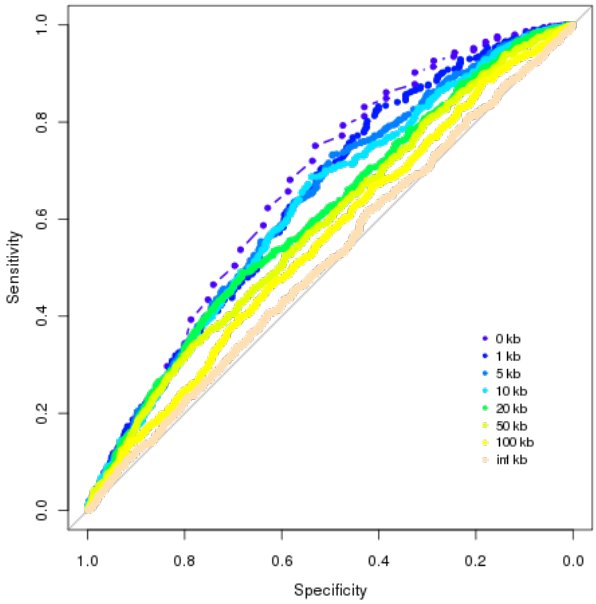
(d) Waffect - statistic S2

Figure 14: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.2-0.3]$, recessive model, $RR1 = 1.0$, $RR2 = 1.6$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

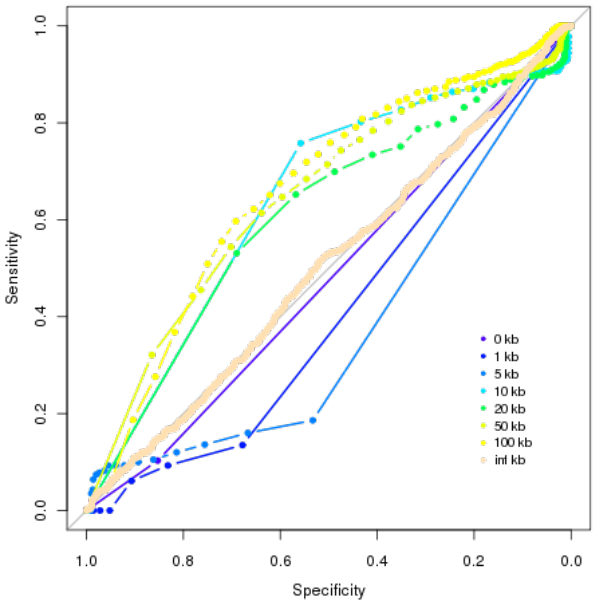
5.2 Simulations for MAF in $[0.2-0.3]$, recessive model, $RR1 = 1.0$, $RR2 = 1.3$



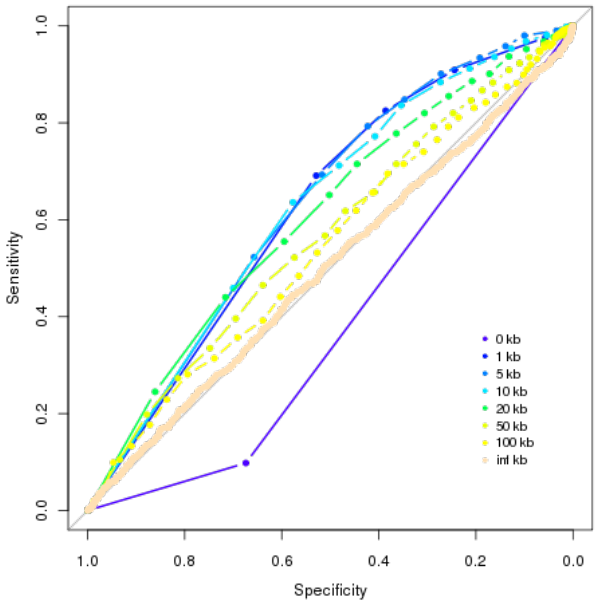
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



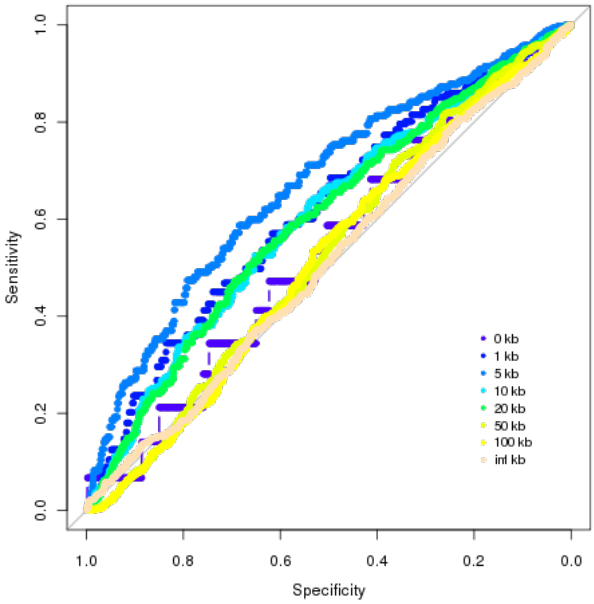
(c) Hapgen - statistic S2



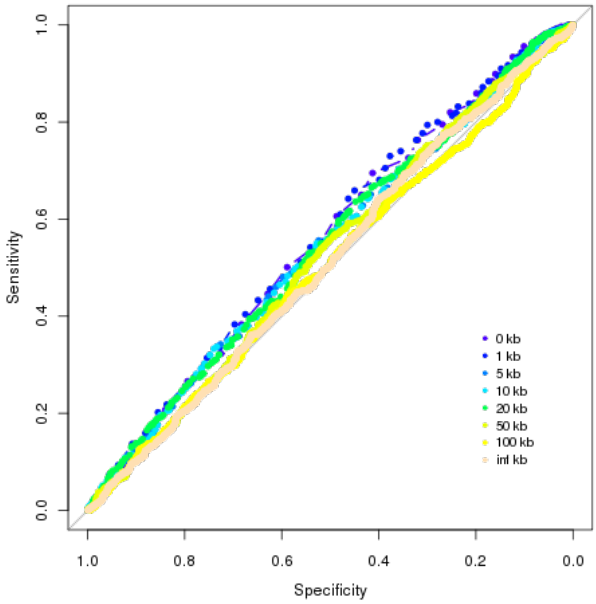
(d) Waffect - statistic S2

Figure 15: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.2-0.3]$, recessive model, $RR1 = 1.0$, $RR2 = 1.3$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

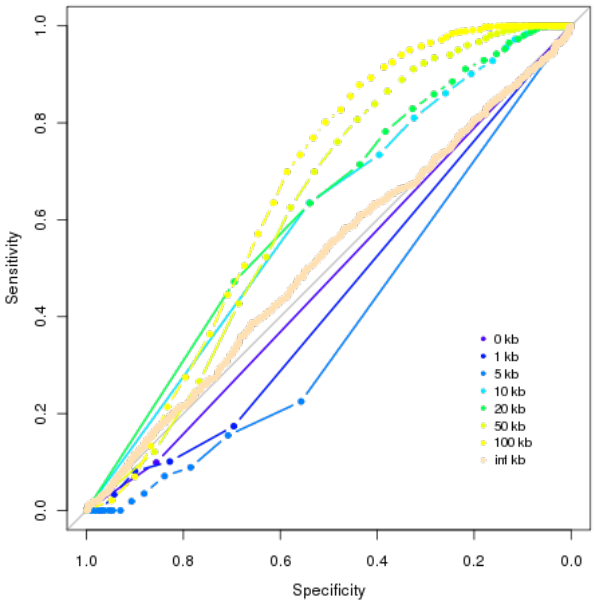
5.3 Simulations for MAF in $[0.1-0.2]$, recessive model, $RR1 = 1.0$, $RR2 = 1.6$



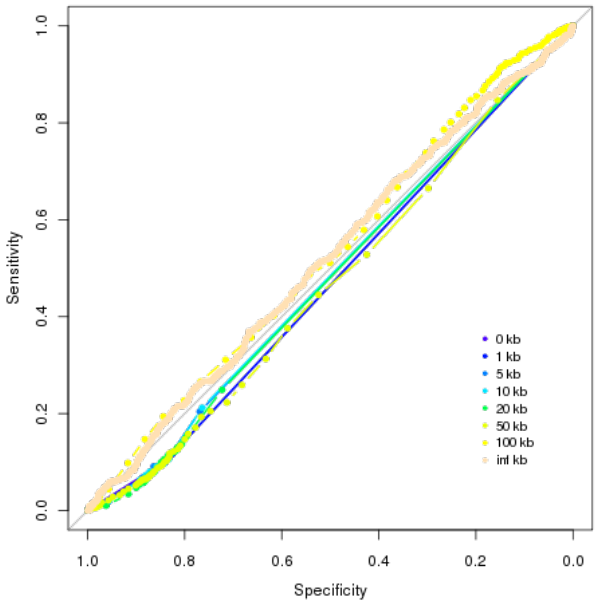
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



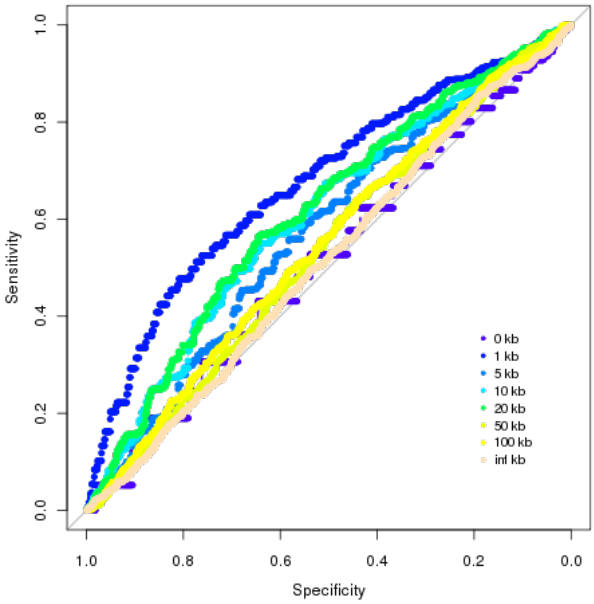
(c) Hapgen - statistic S2



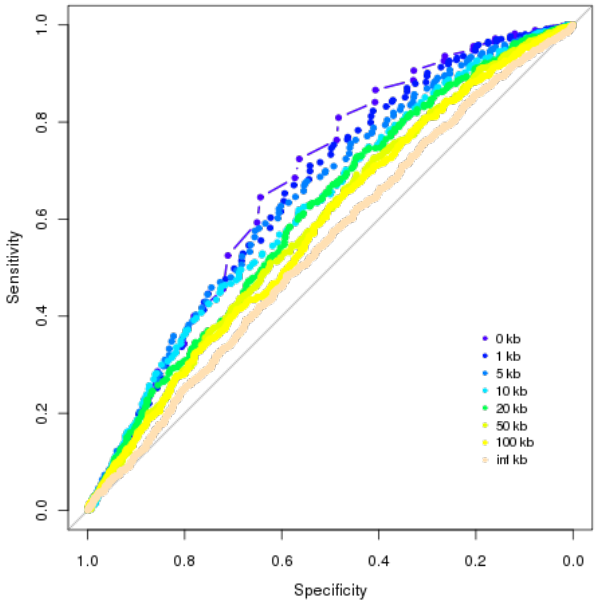
(d) Waffect - statistic S2

Figure 16: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.1-0.2]$, recessive model, $RR1 = 1.0$, $RR2 = 1.6$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

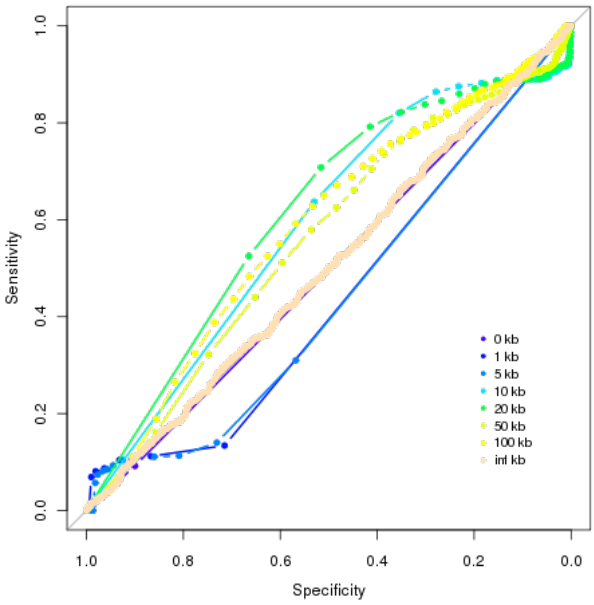
5.4 Simulations for MAF in $[0.1-0.2]$, recessive model, $RR1 = 1.0$, $RR2 = 1.3$



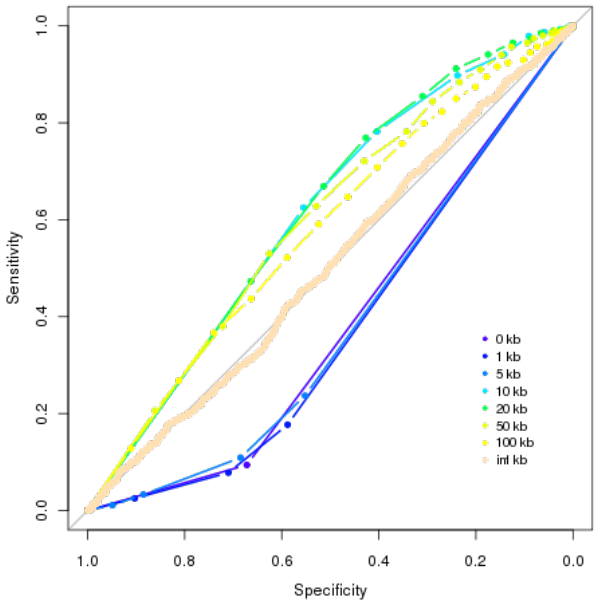
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



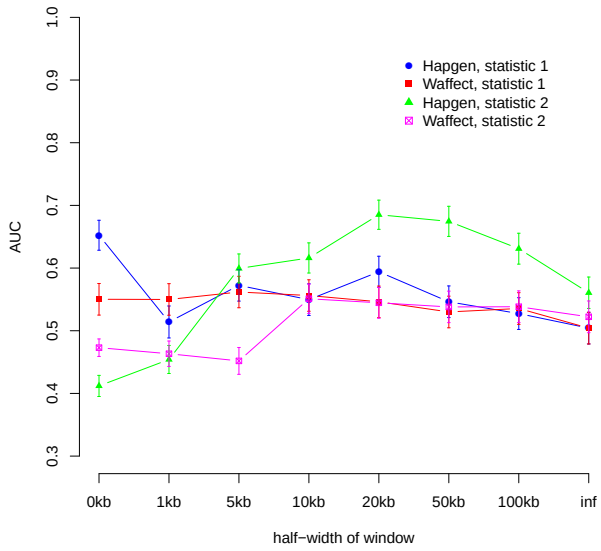
(c) Hapgen - statistic S2



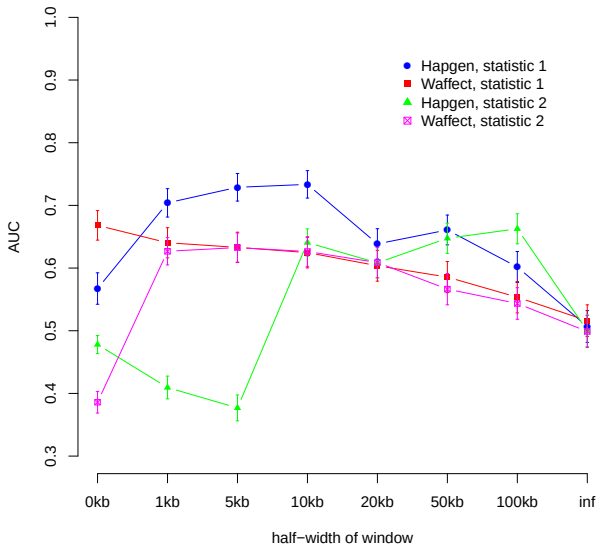
(d) Waffect - statistic S2

Figure 17: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAFin[0.1-0.2]$, recessive model, $RR1 = 1.0$, $RR2 = 1.3$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

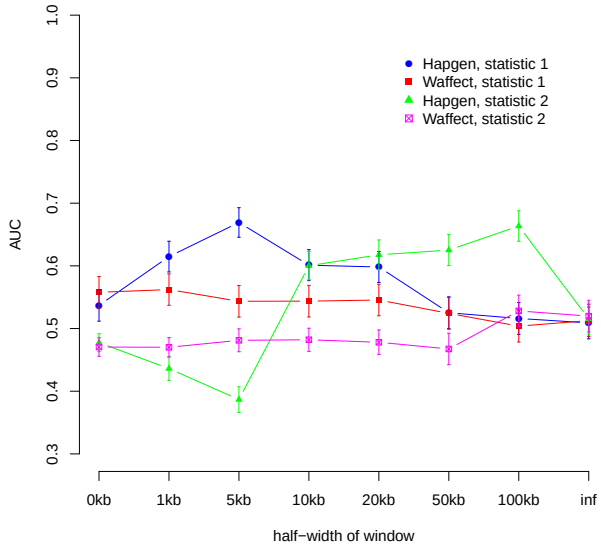
5.5 Recapitulation



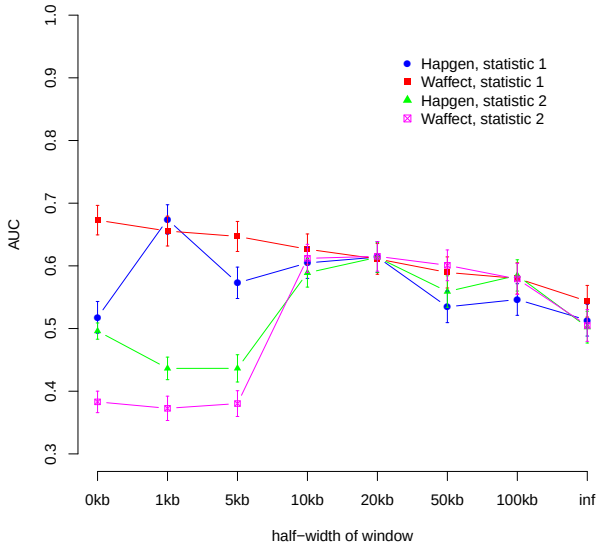
(a) $MAF \in [0.2 - 0.3]$; $RR1 = 1$; $RR2 = 1.6$



(b) $MAF \in [0.2 - 0.3]$; $RR1 = 1$; $RR2 = 1.3$



a) $MAF \in [0.1 - 0.2]$; $RR1 = 1$; $RR2 = 1.6$



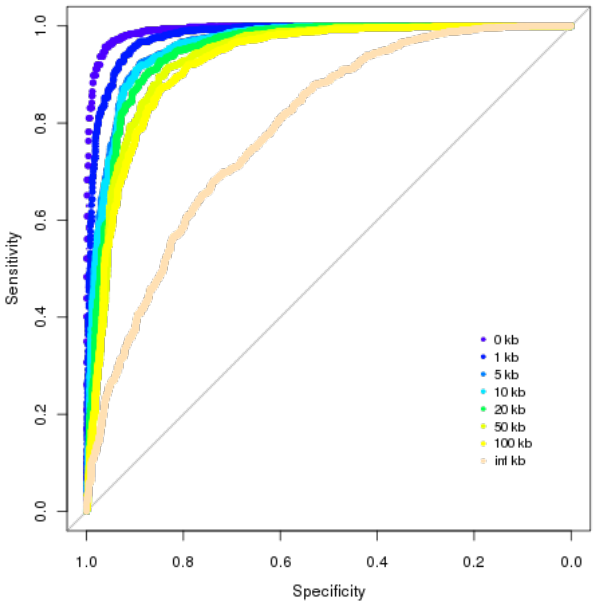
(b) $MAF \in [0.1 - 0.2]$; $RR1 = 1$; $RR2 = 1.3$

Figure 18: Comparison of the area under the curves (AUCs) corresponding to the ROC curves of figures 14, 15, 16 and 17.

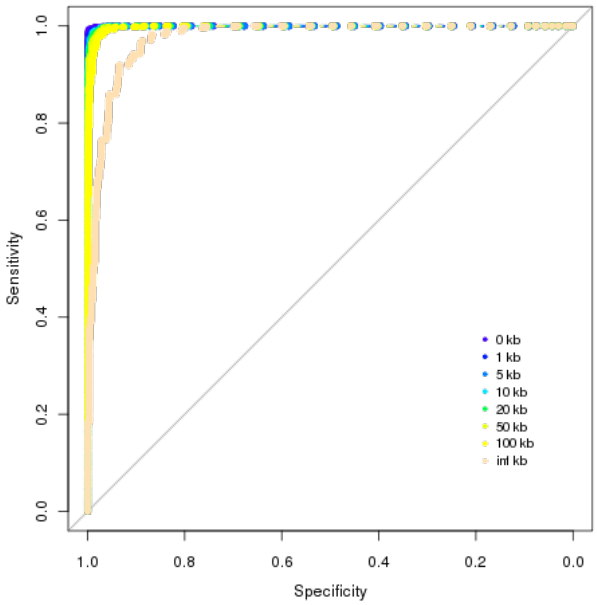
Discussion Identifying statistically significant SNPs is known to be difficult under the recessive hypothesis. Therefore, as expected, similar (bad) trends are observed for both Hapgen and `waffect` methods: AUC values rarely reach 0.7 and the global trend rather sticks to 0.5. Regarding these bad AUC values, is is not worth carrying out a more thorough analysis of the curves.

6 Multiplicative genetic model

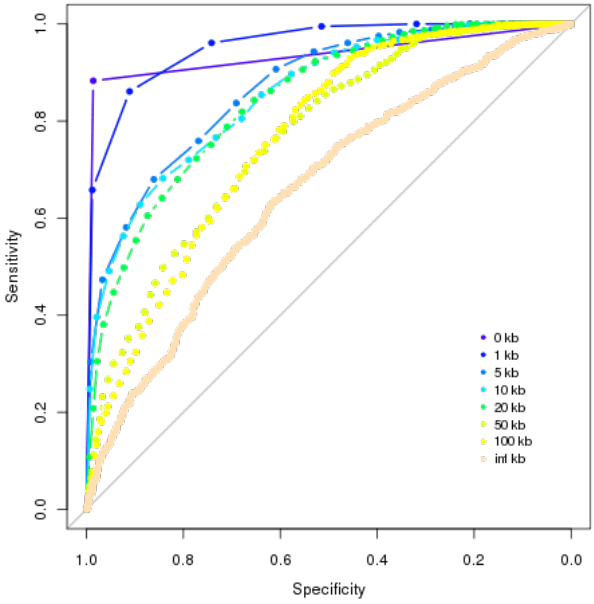
6.1 Simulations for MAF in $[0.2-0.3]$, multiplicative model, $RR1 = 1.6$, $RR2 = 2.56$



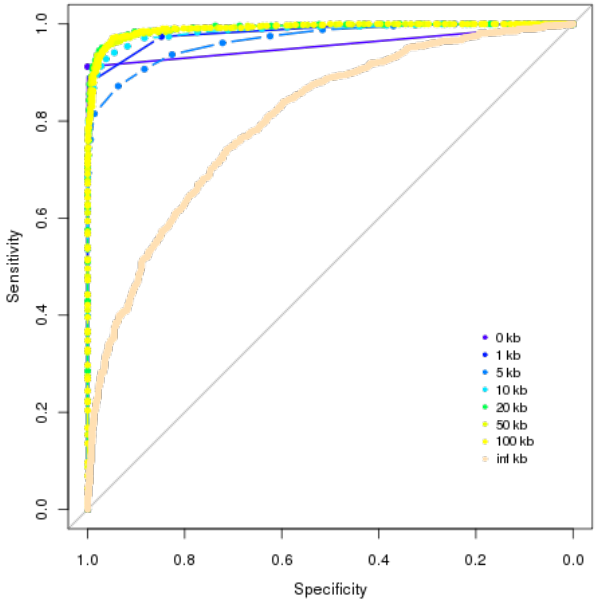
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



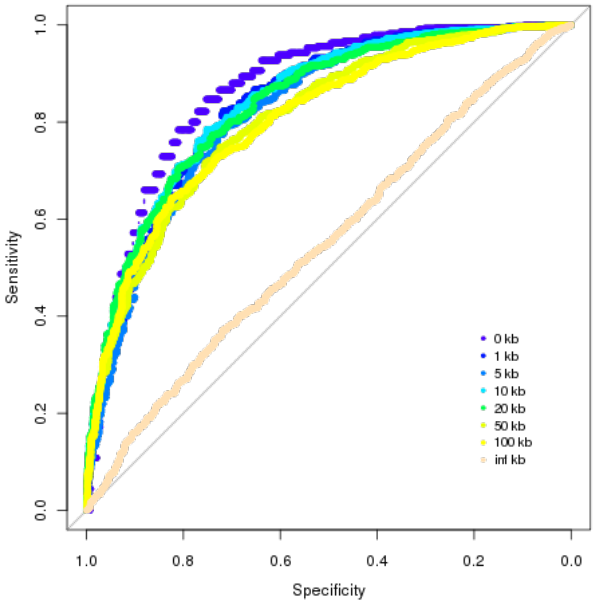
(c) Hapgen - statistic S2



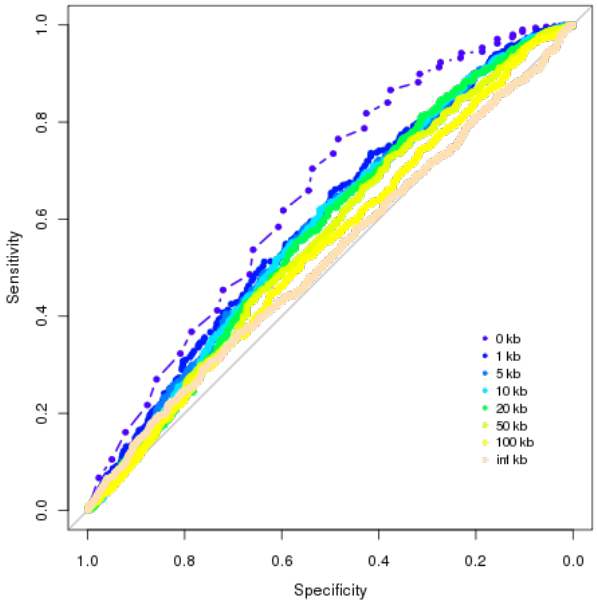
(d) Waffect - statistic S2

Figure 19: Comparison of the ROC curves plotted for 1,000 simulations under H1 and 1,000 simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.2-0.3]$, multiplicative model, $RR1 = 1.6$, $RR2 = 2.56$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

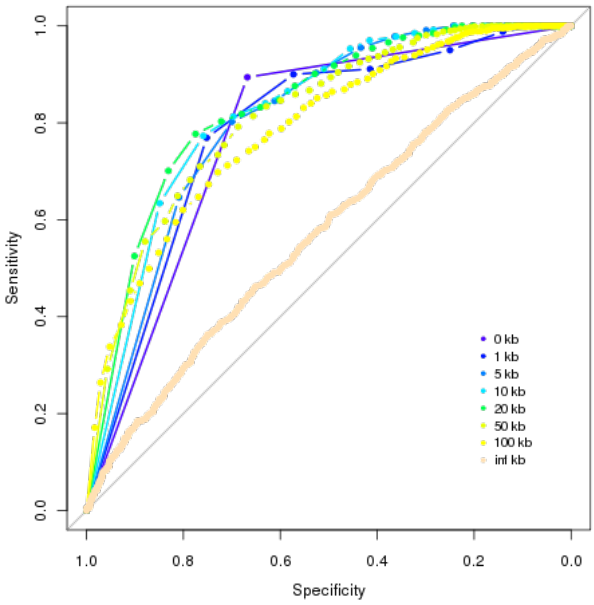
6.2 Simulations for MAF in $[0.2-0.3]$, multiplicative model, $RR1 = 1.3$, $RR2 = 1.69$



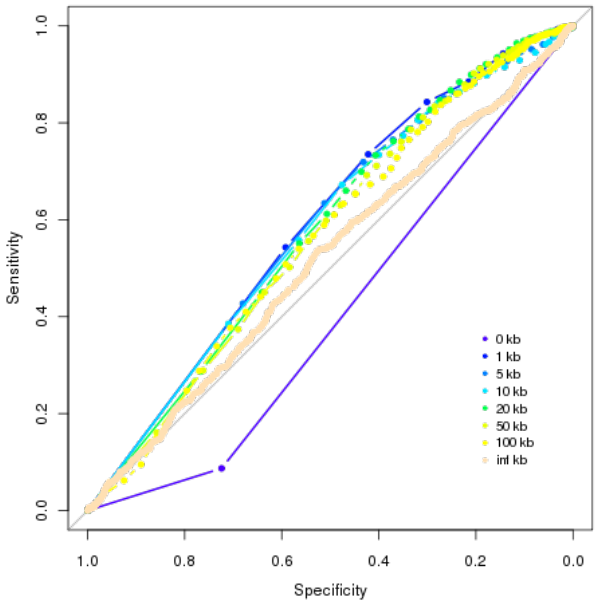
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



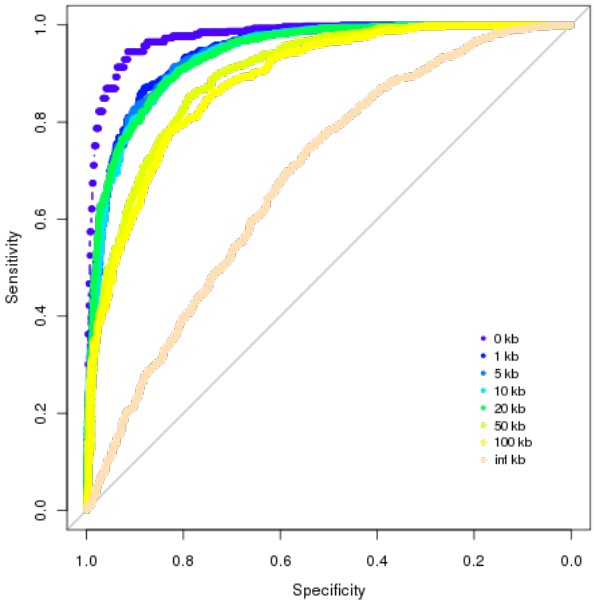
(c) Hapgen - statistic S2



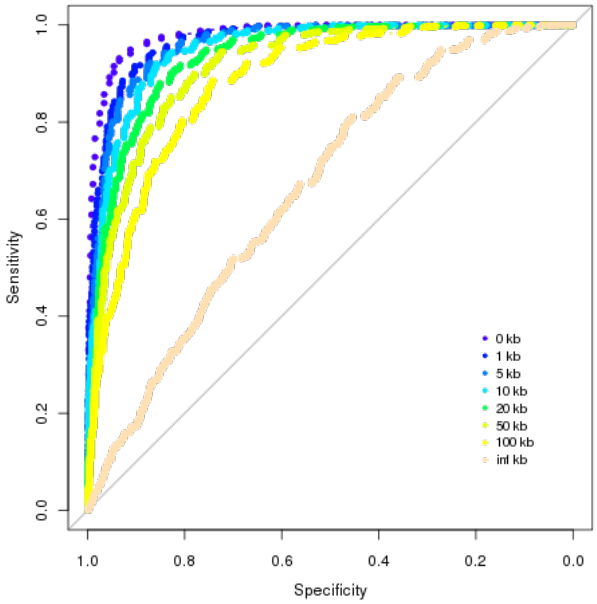
(d) Waffect - statistic S2

Figure 20: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.2-0.3]$, multiplicative model, $RR1 = 1.3$, $RR2 = 1.69$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

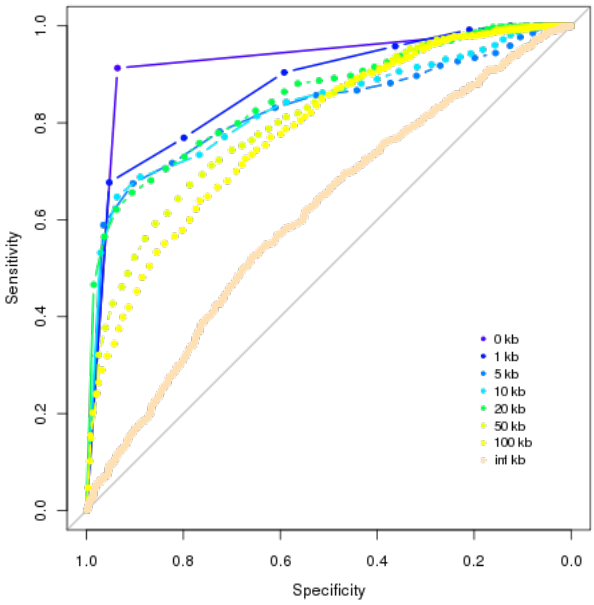
6.3 Simulations for MAF in $[0.1-0.2]$, multiplicative model, $RR1 = 1.6$, $RR2 = 2.56$



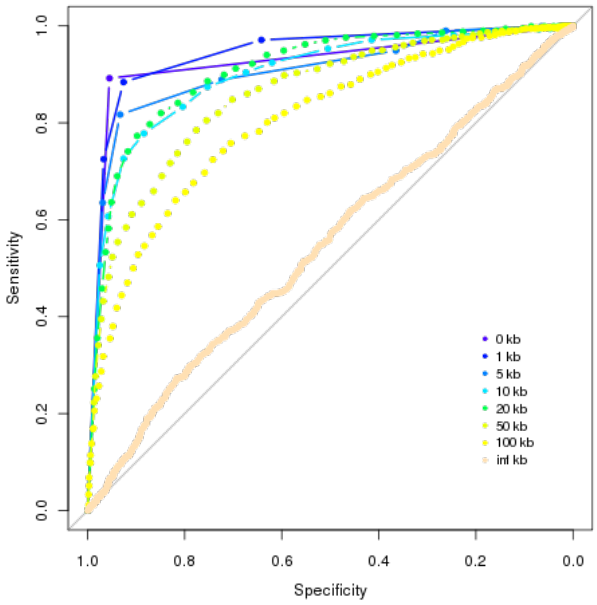
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



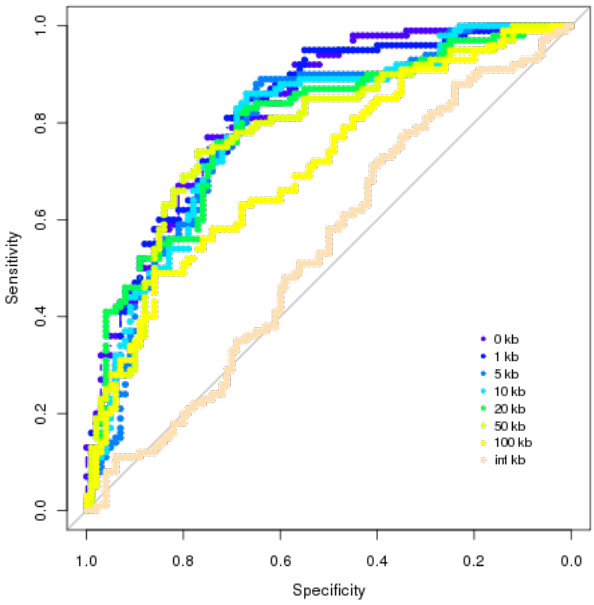
(c) Hapgen - statistic S2



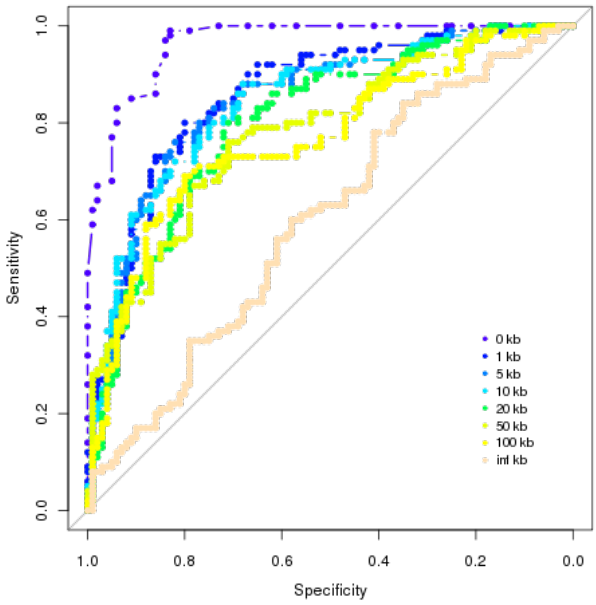
(d) Waffect - statistic S2

Figure 21: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAF \in [0.1-0.2]$, multiplicative model, $RR1 = 1.6$, $RR2 = 2.56$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

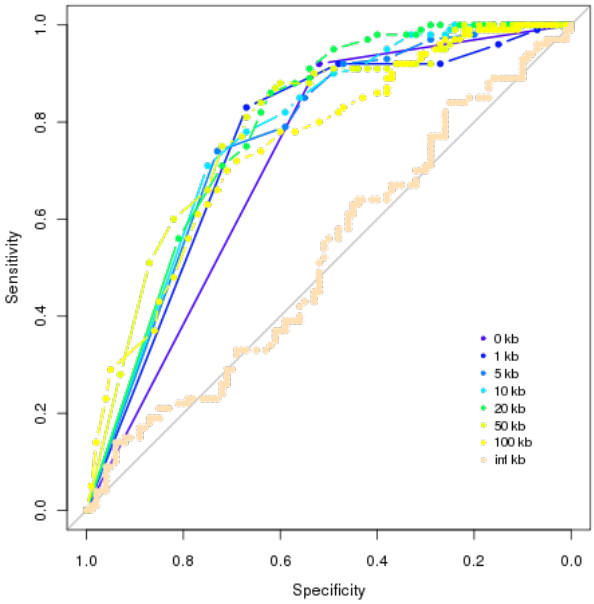
6.4 Simulations for MAF in $[0.1-0.2]$, multiplicative model, $RR1 = 1.3$, $RR2 = 1.69$



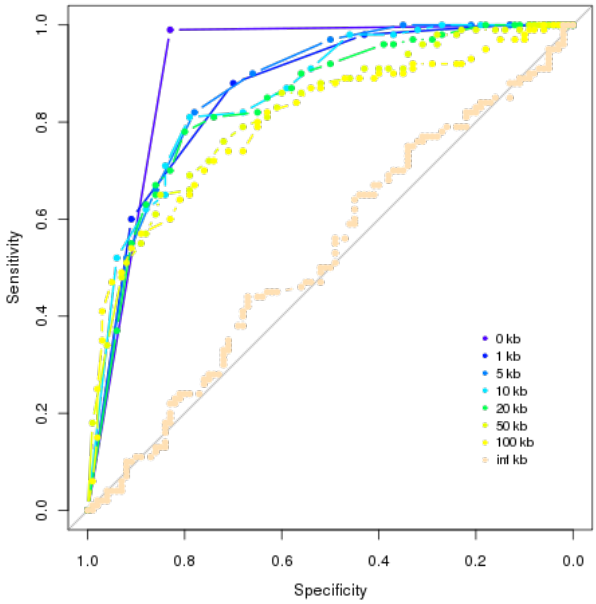
(a) Hapgen - statistic S1



(b) Waffect - statistic S1



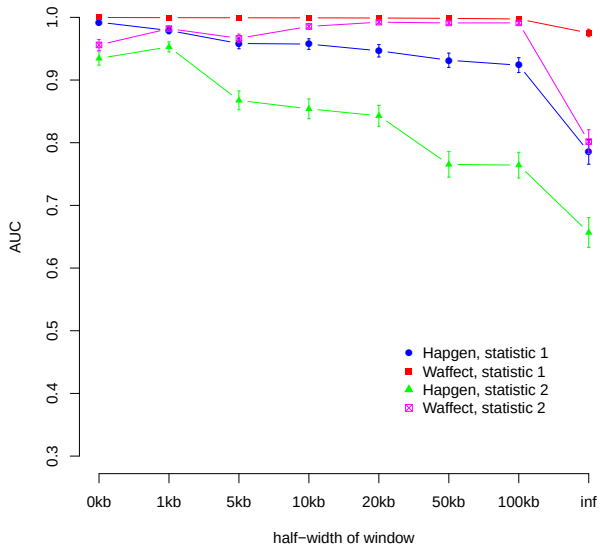
(c) Hapgen - statistic S2



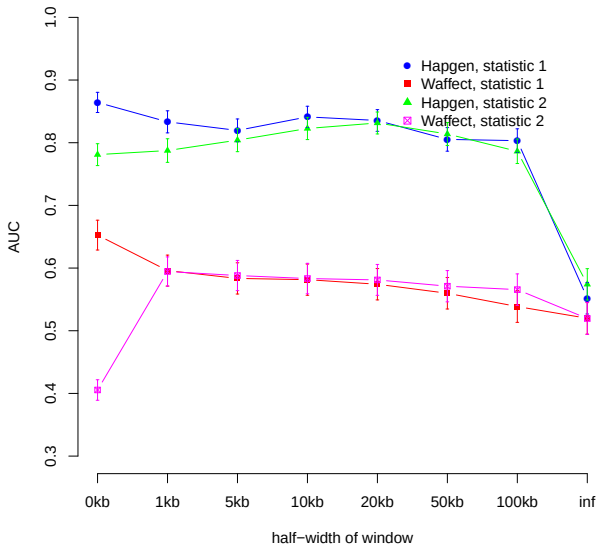
(d) Waffect - statistic S2

Figure 22: Comparison of the ROC curves plotted for $1,000$ simulations under H1 and $1,000$ simulations under H0, respectively, for two statistics S1 and S2 - $MAFin[0.1-0.2]$, multiplicative model, $RR1 = 1.3$, $RR2 = 1.69$. S1 is defined as the maximum $-log_{10}(p-value)$ obtained with the Cochran-Armitage test in a window centered on a reference locus, namely the susceptibility locus under H1. S2 measures the number of single tests with p-values less than or equal to 0.1.

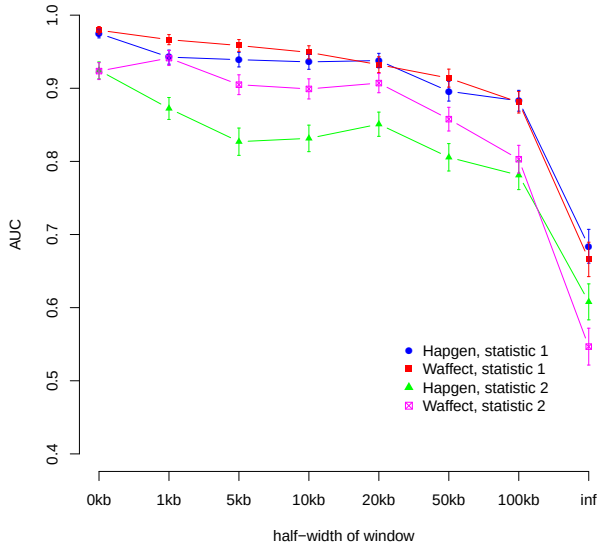
6.5 Recapitulation



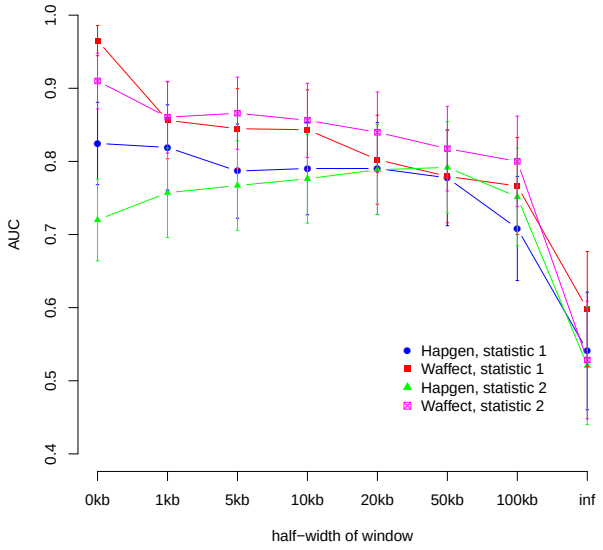
(a) $MAF \in [0.2-0.3]$; $RR1 = 1.6$; $RR2 = 2.56$



(b) $MAF \in [0.2-0.3]$; $RR1 = 1.3$; $RR2 = 1.69$



a) $MAF \in [0.1-0.2]$; $RR1 = 1.6$; $RR2 = 2.56$



(b) $MAF \in [0.1-0.2]$; $RR1 = 1.3$; $RR2 = 1.69$

Figure 23: Comparison of the area under the curves (AUCs) corresponding to the ROC curves of figures 19, 20, 21 and 22.

Discussion In the case of the multiplicative model, similar trends can be observed for Hapgen and `waffect` except in the case of a high MAF and a low risk. In this latter case, whatever the statistic considered, the AUC values roughly differ by 0.2-0.3, in favour of Hapgen. In five out of the six remaining cases (2 statistics $\times$ 3 conditions), the discrimination power is higher for `waffect` . In the case of statistic S1, low MAF and high risk, we can observe that the curves are close. Finally, Hapgen overcomes `waffect` whatever the statistic considered, in the high MAF and low risk case.

References

[1] Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics (International Biometric Society)*, **11**(3), 375–386.

[2] Cochran, W. (1954). Some methods for strengthening the common chi-square tests. *Biometrics (International Biometric Society)*, **10**(4), 417–451.

[3] Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies via imputation of genotypes. *Nature Genetics*, **39**, 906–913.

[4] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., Maller, J., Sklar, P., de Bakker, P., Daly, M., and Sham, P. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, page 81.

[5] Robin, R., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011). proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinformatics*, **12**, 77.

[6] Sonego, P., Kocsor, A., and Pongor, S. (2008). Roc analysis: applications to the classification of biological sequences and 3d structures. *Briefings in Bioinformatics*, **9**(3), 198–209.

[7] Spencer, C., Su, Z., Donnelly, P., and Marchini, J. (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genetics*, **5**(5), e1000477+.