Europe PMC Funders Group
Author Manuscript
Genet Epidemiol. Author manuscript; available in PMC 2011 April 14.

Published in final edited form as:
Genet Epidemiol. 2008 April ; 32(3): 204–214. doi:10.1002/gepi.20295.

Examining the Statistical Properties of Fine-Scale Mapping in Large-Scale Association Studies

Steven Wiltshire[*], **Andrew P. Morris**, and **Eleftheria Zeggini**
Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

Abstract

Interpretation of dense single nucleotide polymorphism (SNP) follow-up of genome-wide association or linkage scan signals can be facilitated by establishing expectation for the behaviour of primary mapping signals upon fine-mapping, under both null and alternative hypotheses. We examined the inferences that can be made regarding the posterior probability of a real genetic effect and considered different disease-mapping strategies and prior probabilities of association. We investigated the impact of the extent of linkage disequilibrium between the disease SNP and the primary analysis signal and the extent to which the disease gene can be physically localised under these scenarios. We found that large increases in significance (>2 orders of magnitude) appear in the exclusive domain of genuine genetic effects, especially in the follow-up of genome-wide association scans or consensus regions from multiple linkage scans. Fine-mapping significant association signals that reside directly under linkage peaks yield little improvement in an already high posterior probability of a real effect. Following fine-mapping, those signals that increase in significance also demonstrate improved localisation. We found local linkage disequiliptium patterns around the primary analysis signal(s) and tagging efficacy of typed markers to play an important role in determining a suitable interval for fine-mapping. Our findings help inform the interpretation and design of dense SNP-mapping follow-up studies, thus facilitating discrimination between a genuine genetic effect and chance fluctuation (false positive).

Keywords

genome-wide association; false positive; localization; disease gene; linkage disequilibrium; haplotype; fine-scale mapping

INTRODUCTION

The search for genes underlying complex traits and diseases enters an exciting new phase, with the publication of genome-wide association (GWA) scans for many complex diseases, including Alzheimer's disease [Coon et al., 2007], breast cancer [Easton et al., 2007], coronary artery disease [Wellcome Trust Case Control Consortium, 2007], Crohn's disease [Hampe et al., 2007; Wellcome Trust Case Control Consortium, 2007], myocardial infarction [Helgadottir et al., 2007] and type 2 diabetes [Sladek et al., 2007; Wellcome Trust Case Control Consortium, 2007; Zeggini et al., 2007]. The GWA scan approach offers genuine prospects for major advances in understanding the genetic aetiology of complex

© 2007 Wiley-Liss, Inc.

[*]Correspondence to: Dr. Steven Wiltshire, Laboratory for Genetic Epidemiology, Western Australian Institute for Medical Research, Queen Elizabeth II Medical Centre, Ground Floor, B Block, Hospital Avenue, Nedlands WA 6009, Australia. shire@cyllene.uwa.edu.au.

The Supplementary materials described in this article can be found at http://www.interscience.wiley.com/jpages/0741-0395/suppmat

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

traits, notwithstanding its difficulties and challenges [Hirschhorn and Daly, 2005]. However, the several hundred complex trait linkage scans published over the course of the past decade or so still provide a wealth of statistical and positional evidence for the presence of susceptibility genes [Wiltshire et al., 2005; Roeder et al., 2006] that could be exploited in a linkage-based fine-scale association mapping approach.

Both experimental approaches are likely to proceed along a design familiar to linkage studies. First, a primary analysis will be conducted with a modestly dense set of markers [such as the commercially available 300 K and 500 K single nucleotide polymorphism (SNP) chip products; Barrett and Cardon, 2006]. Second, potentially interesting regions (i.e. those reaching a certain level of significance) will be investigated in additional samples and/or populations in the hope of replication [Thomas et al., 2004; Skol et al., 2006; Wang et al., 2006]. In parallel, such regions identified in primary analyses will undergo fine-scale mapping with a denser set of markers (e.g. HapMap phase II or resequencing) to obtain better (hopefully stronger) evidence for a genetic effect and better (hopefully more precise) localisation of the disease variant.

Here, we explore the behaviour of primary mapping signals upon fine-mapping, under both null and alternative hypotheses, and examine the inferences that can be made regarding the posterior probability of there being a real genetic effect. We consider several different disease mapping strategies and prior probabilities of association. We examine the impact of regional linkage disequilibtium (LD) architecture and of the extent of LD between the disease SNP and the primary analysis signal — to model different tagging efficacies — on the posterior probability of a real genetic effect. Finally, we consider the extent to which the disease gene can be physically localised under these scenarios.

METHODS

Using a coalescent simulator [Schaffner et al., 2005], we simulated 1,000 separate populations of 10,000 chromosomes each of 500 kb in length (the size being dictated by computational limitations), assuming a European ancestry and population history. For each of these populations, the SNP density was thinned, as necessary, to generate mean SNP separations close to those of the 10 ENCODE (build 16.1c) regions [International HapMap Consortium, 2005]. These 1,000 populations of 10,000 thinned haplotypes constituted the substrates for all subsequent analyses.

For this study, we considered a range of disease models, encompassing three gene effect sizes (minor allele homozygote genotype relative risk (GRR_{hom}) of 1.5, 2 and 3) under three modes of action (dominant, recessive and additive). We also considered the null hypothesis of no genetic effect. For each population, using its set of thinned chromosomes described above, we generated case/control replicates comprising the genotypes of 1,000 control subjects and 1,000 case subjects. A single SNP was designated as the causal variant, selected at random from all polymorphisms with a predetermined minor allele frequency (“close” to 0.1, 0.2 and 0.4). Individuals were generated by selecting pairs of haplotypes at random, with replacement, from the population, and simulating their disease status according to the disease model. The process was repeated until sufficient cases and controls were generated. We retained the position of the disease SNP for each case/control replicate. We repeated our simulations and analyses using solely common (>5%) SNPs.

For each population, for the primary analysis of each case/control replicate, we selected a subset of SNPs from the fine-scale map to achieve a HapMap phase I density of approximately one common (≥5%) SNP every 4 kb and approximately one rare (≤5%) SNP every 30 kb. Each replicate was analysed using logistic regression implemented in

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

COCAphase [Dudbridge, 2003], either in a SNP-by-SNP approach or in a five-SNP haplotype sliding window approach. The most significantly associated SNP, or five-SNP haplotype, together with its position, was recorded. Rare (<1%) haplotypes were pooled during analysis. We continued to simulate and analyse case/control replicates until a significant primary analysis signal — that is one with asymptotic $P \leq 0.001$ (uncorrected) — was achieved. (This significance threshold was chosen for reasons of computational practicality, especially under the null hypothesis, given a simulation study of this nature.) The number of replicates necessary to achieve this (N_{rep}), together with the position of the associated SNP, was recorded. The significant replicate was then fine-mapped by introducing into the analysis map all SNPs (from the thinned haplotypes) that lay within 30 kb on either side of the significantly associated SNP or associated five-SNP haplotype. The size of this region corresponds to approximately three average LD blocks [International HapMap Consortium, 2005], and as such is reasonably well dimensioned for modelling in a simulation study such as this. The fine-mapped region was reanalysed with COCAphase: the most significantly associated SNP (or haplotype), its position and *P*-value recorded. We determined the magnitude and direction of the *P*-value change during fine-scale mapping, and the extent of LD between the primary "hit" SNP and the actual disease SNP from the control sample genotypes.

This exercise was performed for all 1,000 populations under the null model (of no genetic effect) and three alternative genetic models, for each genetic effect size and allele frequency. For each set of parameters, we were, therefore, able to derive two measures: the total number of replicates necessary to achieve a significant primary analysis signals meeting the significance threshold for all 1,000 populations (N_{rep}), and the number of primary analysis signals that increased in significance following fine-scale mapping (N_{inc}). From these, we could calculate the following quantities:

a. $P(s|H_0)$, $P(s|H_1)$ — the probabilities of obtaining a significant result during primary mapping, computed as $N_{rep}/1{,}000$, under the null and alternative hypotheses, respectively.

b. $P(i|s,H_0)$, $P(i|s,H_1)$ — the conditional probabilities of observing an increase in statistical significance of an association following fine-mapping, computed as $N_{inc}/1{,}000$, under the null and alternative hypotheses, respectively.

c. $P(nd|s,H_0)$, $P(nd|s,H_1)$ — the conditional probabilities of observing either no change or a decrease in the significance of an association following fine-mapping, under the null and alternative hypotheses, calculated as $1-P(i|s,H_0)$ and $1-P(i|s,H_1)$, respectively.

The prior probability of a SNP in the genome being pathologically important will depend not only on the genetic architecture of the trait (given the total number of aetiologically important genes, any interactions between them and their physical distribution along the chromosome) but also on the experimental scenario (contrast a de novo scan with the association mapping of a linkage peak). Therefore, we have decided upon three examples for the purposes of our study to encompass a realistic range of values for the experimental and biological scenarios described above. Each $P(H_1)$ is the prior probability of a single disease SNP lying on the 500 kb DNA segment considered in our analyses. The three scenarios are as follows:

1. Ten disease-causing SNPs in the genome (of 3,000 Mb), investigated as part of a de novo GWA scan, each of which has an equal probability of lying on our 500 kb DNA segment: $P(H_1) = 10 \times (\text{segment length})/(\text{genome length})\ [(\text{segment length})/(\text{genome length})]^9 = 0.00166$.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

2. A single disease-causing SNP lying somewhere within a 20 Mb (~20-cM) linkage peak that has been replicated by multiple genome-wide linkage studies: $P(H_1) = 5 \times 10^5/2 \times 10^7 = 0.025$.

3. A single disease-causing SNP in the genome, investigated as part of a typical (10 cM) genome-wide multipoint linkage scan of a late-onset trait such as type 2 diabetes, with an allele-sharing LOD score of 3 or more: $P(H_1) = 0.667$ (from [Wiltshire et al., 2005]).

Given these three prior probabilities for a disease SNP lying on our 500 kb DNA segment ($P(H_1)$), the corresponding prior probability for no disease SNP ($P(H_0) = 1–P(H_1)$), and the quantities in (a), (b) and (c), we used Bayes' theorem to obtain the following posterior probabilities:

(d) The posterior probability of a significant primary map association being due to a genuine disease gene before fine-mapping: $P(H_1|s) = P(s|H_1)P(H_1)/[P(s|H_1)P(H_1)+P(s|H_0)P(H_0)]$.

(e) The posterior probability of a significant primary map association being due to a genuine disease gene after fine-mapping, given that it increases in statistical significance (i.e. the *P*-value gets smaller): $P(H_1|i,s) = P(i|s,H_1)P(s|H_1)P(H_1)/[P(i|s, H_1)P(s|H_1)P(H_1)+P(i|s, H_0)P(s|H_0)P(H_0)]$.

(f) The analogous posterior probability, $P(H_1|nd, s)$, given no change or a decrease in statistical significance of the association following fine-mapping.

Many (perhaps most) fine-mapping association studies will be conducted using tag SNPs, selected in order to reduce genotyping effort by exploiting the LD patterns in the genome [Carlson et al., 2004; de Bakker et al., 2006]. We have, therefore, examined the effect of the LD between our marker SNPs (standing in for tags) and the disease gene (present somewhere with a given prior probability) by stratifying our results according to three levels of LD between the disease gene and the primary analysis signal: low ($r^2 < 0.5$, in which the disease gene is poorly tagged), moderate to high ($r^2 \geq 0.5$) and high ($r^2 \geq 0.8$, in which the disease gene is well tagged). The quantities (a), (b) and (c) above are now conditional on r^2, and under the alternative hypothesis of a genuine genetic effect, $P(s|H_1, r^2)$ and $P(i|s, H_1, r^2)$, are recalculated solely from the replicates meeting the r^2 criterion; under the null hypothesis, r^2 is irrelevant (there is no disease gene) and $P(s|H_0, r^2)$ and $P(i|s,H_0, r^2)$ will, therefore, be the same as in their unconditional counterparts. The posterior probabilities of a real gene conditional upon r^2, both before and after fine-mapping — $P(H_1 |s, r^2)$ and $P(H_1 |i,s, r^2)$ — are calculated as before from these quantities.

RESULTS

We consider here a disease gene with an allele frequency of 0.2 with several different effect sizes acting under dominant, recessive and additive genetic models. The probabilities of obtaining a significant association (i.e. $P<0.001$, uncorrected), and of observing an increase in statistical significance after fine-mapping, for both individual SNP and five-SNP haplotype sliding windows are shown in Table I. Under the null hypothesis of no genetic effect, we see a primary analysis signal increase in significance by approximately one third of the time in a single-SNP and five-SNP haplotype-based analyses following fine-mapping (Table II). Only in a minority of these instances (5% for single SNP and 11% for haplotype-based) do the *P*-values change by one or more orders of magnitude. We never see an increase in statistical significance of three or more orders of magnitude during single-SNP or haplotype-based analyses (Table II).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Under the alternative hypothesis of a genetic effect, increases in the primary analysis signal significance upon fine-mapping are much more frequent under dominant and additive models than under the null hypothesis (Table I). This is especially so for large increases (of two or three orders of magnitude) (Table II), and there is little difference between the single SNP and haplotype analytical approaches with significance changes of this order of magnitude. However, irrespective of the analytical approach, recessive genes with small effect sizes barely differ from the null model in terms of power to detect the initial signal (P(s)) the frequency of signal increase on fine-mapping (P(i|s)), or the magnitude of the change in significance, and differ only modestly for larger (GRR_{hom}>2) gene effect sizes (Tables I and II).

THE POSTERIOR PROBABILITY OF A GENUINE GENETIC EFFECT

Using these simulation results, we determined the posterior probability of a real disease gene for three mapping scenarios described above, both before fine-mapping ($P(H_1|s)$) and after fine-mapping ($P(H_1|i,s)$ given an increase in statistical significance. In the first, we consider a GWA scan of a trait with 10 causative genes, each of equal effect and with the same probability of lying on our 500 kb region of DNA (Table III). The prior probability of one of these genes lying within such a region is small, at 0.00167. During a single-SNP analysis of dominant and additive genetic models, the proportional gain in the posterior probability of a genetic effect (i.e. contrasting the primary mapping with fine-scale mapping) is moderate to large; for recessive genes, it is much smaller. For example, fine-mapping an additive gene (with GRR_{hom} of 1.5, 2 or 3), which has been detected in the primary association scan, results in a proportional gain of 60–110% in the posterior probability of a real effect, given an increase (of any size) in the statistical significance of the primary scan signal following the fine-mapping; however, the actual posterior probabilities themselves remain small, between 0.01 and 0.04. During haplotype analyses of common and rare SNPs (Table IV), the proportional increases in posterior probability of a genetic effect are smaller than those under single-SNP analyses, but the posterior probabilities themselves are higher; the converse is true, however, when focusing solely on common SNPs, where both quantities are larger (see supplementary online information).

In the second scenario, we consider a single disease SNP lying somewhere under a 20-Mb linkage peak, with prior probability of 0.025 (Table III). Increases in the posterior probability of an additive or dominant gene, when the primary analysis association signal increases in significance following fine-mapping, are modest and smaller than in the previous scenario, but otherwise the trend is the same. Recessive genes respond the least well to fine-mapping: even with the largest gene effect size (GRR_{hom} = 3), they reach a posterior probability of only 0.123 given after fine-mapping in a single-SNP analysis, whereas under additive and dominant models, the posterior probabilities exceed 0.4. A change in analytical method (i.e. the use of haplotype-based analyses; Table IV) results in the same direction of effect as seen with the genome-wide mapping scenario discussed above, although to a lesser extent.

Our third scenario assumes a single-SNP at the top of a linkage peak observed with a LOD score of 3 or more (Table III). In this instance the prior probability of association is 0.667 [Wiltshire et al., 2005]. The posterior probabilities before fine-mapping are all high (0.8 to over 0.9 for dominant and additive genes; over 0.7 for recessive genes) and the proportional changes given an increase in significance following fine-mapping are negligible, as are the differences between single SNP and haplotype analyses (Table IV).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

THE MAGNITUDE OF THE INCREASE IN SIGNAL SIGNIFICANCE

We examined the changes in the posterior probability of a genetic effect in terms of the magnitude of the increase in statistical significance of the association signal seen in two of these mapping scenarios — 10 SNPs somewhere in the genome, and one SNP somewhere under a 20-Mb linkage peak (Table V). It is clear from Table II that increases of three orders of magnitude in single-SNP and haplotype-based analyses are never seen under the null hypothesis of no gene. Consequently, if fine-mapping results in an increase of this magnitude, the posterior probability of a genetic effect will be one. However, increases of this size are uncommon (except for large gene effect sizes under dominant models — unlikely in complex traits) and we, therefore, focus on the more common and practically useful increase of at least one order of magnitude. Increases of this size have a pronounced effect on the posterior probability of a genetic effect (Table V). For example, for our three additive models (GRR_{hom} of 1.5, 2.0 and 3.0), the posterior probabilities of a genetic effect given an increase in significance of at least one order of magnitude after fine-mapping are 0.033, 0.244 and 0.423, respectively, (increases of 378, 1327 and 2023%) for the scenario of 10 genes in the genome (Table V). (Dominant genes show bigger proportional changes, whereas recessive genes respond only modestly.) For the scenario of a single gene under a 20-Mb linkage peak, these probabilities are 0.347, 0.832 and 0.919 for the three additive genetic models described above, with more modest proportional changes of 255–285%. (Recessive genes show a response comparable with additive and dominant genes only with large effect sizes under this mapping scenario; Table V.)

THE EFFECT OF LINKAGE DISEQUILIBRIUM BETWEEN THE MARKER AND DISEASE SNPS

We considered the effects of the LD between the disease SNP and marker SNP most strongly associated during primary analysis with the response to fine-mapping in the same two scenarios as those above. In the case of dominant genes, the greatest increases in the posterior probability of a genetic effect given an increase in significance tend to occur when the LD between disease and marker SNP is low (r^2<0.5), and the smallest proportional increases are seen when the LD is high, irrespective of marker allele frequency, although for small gene effect sizes, the difference is not large (Table VI). Similar findings are seen for larger gene effect sizes under additive models: for example, for an additive gene with GRR_{hom} = 2.0, we see proportional increases of 130% with a low r^2, 93% with a moderate r^2 (≥0.5) and 75% with a high r^2 (≥0.8), in the genome-wide mapping scenario (Table VI). The same is true when fine-mapping a single SNP under a well-replicated linkage peak. However, the converse is seen for recessive genes, with the greatest gains in posterior probability to be made when the LD between disease and marker SNPs is moderate or high. For the same gene effect size (GRR_{hom} = 2.0), we see only a 27% gain for the low r^2 stratum, contrasting with gains of 90% and 71% for the moderate and high r^2 strata, respectively, in a genome-wide mapping scenario. A similar trend is seen when fine-mapping recessive genes under well-replicated linkage peaks.

CONSEQUENCES OF NO CHANGE IN SIGNIFICANCE FOLLOWING FINE MAPPING

We examined the consequences of no change in the statistical significance of association (for single SNPs) or a decrease (for haplotype analyses) on the posterior probability of a genetic effect ($P(H_1|nd, s)$ for our three experimental fine-mapping scenarios (Tables VII and VIII). In each case, the posterior probabilities fall given no change/a decrease in the significance of the primary analysis result, more so for haplotype analyses than for single

SNP analyses (Tables VII and VIII). This decrease is negligible when fine-mapping a linkage peak with LOD≥3, and is only modest for single-SNP analyses of dominant and additive genes (negligible for recessive genes) in our first two fine-mapping scenarios, irrespective of marker SNP frequency.

EFFECTS OF MARKER AND DISEASE ALLELE FREQUENCIES ON THE RESPONSE TO FINE-MAPPING

We examined the effects of marker and disease allele frequencies on the response to fine-mapping in additional simulations. These results are shown in the online supplementary material. In general, the use of common marker SNPs (frequency >0.05) in the primary and fine-scale mapping process yields the same pattern of response as the use of common and rare SNPs, although the magnitude of the responses tends to be smaller. In models with a higher disease allele frequency (of 0.4), the frequency of an increase in significance following fine-mapping is only modestly higher, as are the posterior probabilities of a genetic effect (both before and after such an increase), and the proportion gains therein, although, excepting in the case of recessive genes, which showed a markedly improved response to fine-mapping. For models with the lower disease allele frequency (of 0.1) these quantities are smaller, especially so in the case of recessive genes, which respond poorly to fine mapping.

PHYSICAL LOCALISATION OF THE DISEASE GENE

Our simulations provide some insight into the ability to physically localise disease SNPs during a primary analysis (Tables I and IX). Recessive genes are localised poorly during primary analysis, with a mean separation of 147 kb between the disease gene and the primary analysis signal; dominant genes are localised the best, with a mean separation of 48 kb from the primary marker SNP (Table I). However, localisation improves as the r^2 between disease and marker increases (Table IX). For example, an additive gene with $GRR_{hom} = 2$ is localised with a mean separation of 80 kB when the r^2 is low, 23 kb with moderate-to-high LD ($r^2 \geq 0.5$); the very best tags ($r^2 \geq 0.8$) localise the disease gene best, with a mean separation of 15 kb. Primary analysis signals that subsequently increase in significance following fine-mapping are closer to the disease gene than those that remain unchanged (Table IX): for the genetic model adduced above, the mean separations are 58, 22 and 16 kb, for low, moderate and high r^2 strata, respectively. Following fine-mapping, those signals that increase in significance show improved localisation, with mean separations of 52, 18 and 14 kb for low, moderate and high r^2 strata, respectively (data not shown). We note, however, that our measures of physical localisation — especially for those in the low r^2 stratum — will be underestimated owing to the size of the fine-mapping interval we used in this study. We see the same trends in physical localisation irrespective of disease or marker SNP frequency, although, as expected, localisation improves as disease frequency increases (data not shown).

DISCUSSION

There are four principal factors that influence the response to fine-mapping, from which useful inferences can be made when deciding on follow-up studies. The first of these is the underlying genetic model of the trait in question. The ability to discern between a genuine genetic effect and chance fluctuation (false positive) derives from the differences in the distribution of the test statistic under the two hypotheses: a central χ^2 under the null, and a non-central χ^2 under the alternative hypothesis of a genuine genetic effect, in which the non-centrality parameter is related to the size of this genetic effect. Consequently, although false positive signals increased quite noticeably when-fine mapped, they do so less

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

frequently, and by smaller amounts, than those due to a genuine genetic effect. The larger the genetic effect and the higher its allele frequency, the wider this distinction becomes. As might be expected, dominant genes show the greatest frequency (and magnitude) of increases, whereas recessive genes show the lowest frequency — these are barely distinguishable from false positives when the disease allele frequency is low. Large increases appear to be the exclusive domain of genuine genetic effects, especially dominant and additive models; observing such an increase in a study should lead to a strong supposition of the association signal as being due to a real gene. Increases of a single order of magnitude or more, however, are common for dominant and additive genes, but rare under the null hypothesis, and as such provide good discerning power between the two hypotheses, yielding substantially increased posterior probability of a real gene. However, the rarity of such events given a recessive mode of inheritance — except when the disease allele frequency is high — underlies the potential difficulties encountered with mapping such traits.

The second factor is the choice of markers and analytical method. On the whole, haplotype-based methods have comparable power to detect a genetic effect, but higher frequency of increases and lower false positive rates than single-SNP analyses. As a result, there are smaller gains in the posterior probability of a genuine genetic effect given an increase in statistical significance after fine-mapping, although the posterior probabilities themselves are higher, than in the single-SNP analysis. The discerning power of a haplotype-based test may therefore be less than that of a single-SNP analysis, but the overall evidence it provides in terms of the probability of a real effect is greater. Consequently, a well-designed analysis will include both single-SNP and haplotype-based analyses.

The third factor is prior the probability of a genetic effect. When the prior probability is high, such as that typical of a study reporting very strong evidence for linkage with narrow 1-LOD support intervals, the posterior probability of association is also high (given a significant primary association result) before fine-mapping, and the gains given an increase in statistical significance (of any size) are minimal, regardless of the genetic effect size or disease or marker allele frequency. In this circumstance, there is little to gain from the exercise of fine-mapping, and an investigator would do better to direct resources and effort into replication studies from the outset. When the prior is lower — for instance, a broad but well-replicated linkage peak, or the signals from a GWA scan of a complex trait with multiple underlying genes, there can be sizeable increases in the posterior probability of a genuine genetic effect. These gains are larger for dominant and additive genetic models and for genome-wide scans than for linkage peaks. Recessive genes show the least response and will present a particular experimental challenge to map when the disease allele frequency is low. Our study, however, is agnostic to the underlying distribution of functional DNA elements; in a real data situation, the prior probability of a disease-causing mutation would be higher over known functional elements — exons, splice sites, regulatory regions and evolutionarily conserved elements [Hirschhorn and Daly, 2005; Rigoutsos et al., 2006]. These should clearly be accounted for in any posterior probability calculations involving real data.

The fourth factor is the extent of LD between the disease and marker SNPs. The extent of disequilibrium, the allele frequencies at both loci and their phase — encapsulated by r^2 — are powerful determinants of the ability to detect a genetic effect on a trait [Zondervan and Cardon, 2004]. We did not explicitly consider completely random subsets of SNPs (to model use of the Affymetrix 500 k chip) or specifically chosen tag SNPs (to model use of the Illumina 550 k chip) as to do so would have been computationally unfeasible in a simulation study such as this. Nevertheless, our findings will be applicable to analyses with such SNP chips without too much difficulty as we are concerned here principally with the probabilities

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

of association and not the genotyping effort itself. Our study shows that the greatest gains in the posterior probability of a dominant or additive disease gene are seen for disease genes that are poorly or moderately well captured by tags during the primary analysis. Conversely, there tends to be the least to gain from fine-mapping when the disease gene is well captured by marker SNPs (regardless of whether the marker set contains rare alleles or not) with r^2>0.8 — as would be the case with a set of efficacious tag SNPs. This is especially true in GWA scans. The situation for recessive genes is almost the complete opposite.

The ability to localise a disease-causing mutation during fine-mapping studies is influenced by the complex relationship between r^2 and the physical distance on the broad and fine scales. Nonetheless, markers in LD with the disease gene with a high r^2 localise it more tightly during primary analysis (approximately the length of one average LD block), than when the correlation between disease and marker SNPs is low, although in both situations the range of physical distances can be large. An investigator needs to consider, therefore, the local LD patterns around their primary analysis signal(s) and the tagging efficacy of their markers, when determining a suitable interval for fine-mapping. If the fine-mapping interval is inadequately sized, the physical localisation will be bad, even if fine-mapping is accompanied by sizeable increases in the posterior probability of a genuine genetic effect. In either case, the lack of a relationship between r^2 and the physical distance on the fine-scale sets an upper limit on the ability of fine-scale mapping to localise disease genes and the investigator is left with little option but a functional assessment of the most significantly associated SNPs after fine-mapping.

In summary, fine-mapping significant, well-localised linkage peaks yields little improvement in an already high posterior probability of a genuine genetic effect, and the investigator is best directed to replication studies. Large increases in peak significance upon fine-mapping signals from GWA scans or consensus regions from multiple linkage scans can lead to sizeable increases in the probability of the signal indicating a genuine genetic effect. This is especially so when the markers used in the primary analysis do not capture the genetic variation in the region well, but such markers do not localise well in the primary analysis, necessitating larger fine-mapping intervals to more precisely localise the disease gene. Primary analyses with a set of markers (or tags) that efficaciously capture the genetic variation yield only modest improvements in what is already a high probability that the signal is genuine; the exception to this is when mapping recessive genes, in which the use of good tags during primary analysis is crucial. The course of action chosen by an investigator therefore depends largely on resources, experimental scenario and prior hypothesis of the aetiology of the genetic disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

S.W., A.P.M. and E.Z. are supported by the Wellcome Trust. We thank Professor Mark McCarthy (Oxford Centre for Diabetes, Endocrinology and Metabolism) for helpful comments on the manuscript.

Contract grant sponsor: Wellcome Trust.

REFERENCES

Barret JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet. 2006; 38:659–662. [PubMed: 16715099]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet. 2004; 74:106–120. [PubMed: 14681826]

Coon KD, Myers AJ, Craig DW, Webster JA, Pearson JV, Lince DH, Zismann VL, Beach TG, Leung D, Bryden L, Halperin RF, Marlowe L, Kaleem M, Walker DG, Ravid R, Heward CB, Rogers J, Papassotiropoulos A, Reiman EM, Hardy J, Stephan DA. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. J Clin Psychiatry. 2007; 68:613–618. [PubMed: 17474819]

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet. 2006; 37:1217–1223. [PubMed: 16244653]

Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. Genet Epidemiol. 2003; 25:115–121. [PubMed: 12916020]

Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, SEARCH collaborators; Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo KY, Noh DY, Ahn SH, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low YL, Bogdanova N, Schurmann P, Dork T, Tollenaar RA, Jacobi CE, Devilee P, Klijn JG, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, MacPherson G, Reed MW, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko YD, Spurdle AB, Beesley J, Chen X, kConFab; AOCS Management Group. Mannermaa A, Kosma VM, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BA. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 28:1087–1093. Online publication 27 May 2007, doi:10.1038/nature05887. [PubMed: 17529967]

Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, De La Vega FM, Briggs J, Gunther S, Prescott NJ, Onnie CM, Hasler R, Sipos B, Folsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, Schreiber S. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. Nat Genet. 2007; 39:207–211. [PubMed: 17200669]

Helgadottir A, Thorleifsson G, Manolescu A, Gretasdottir S, Blondal T, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgeirsson G, Thorsteinsdottir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. Science. 2007; 316:1491–1493. Published online 3 May 2007. 10.1126/science.1142842. [PubMed: 17478679]

Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Genet Rev. 2005; 6:95–108.

International HapMap Consortium. A haplotype map of the human genome. Nature. 2005; 437:1299–1320. [PubMed: 16255080]

Rigoutsos I, Huynh T, Miranda K, Tsirigos A, McHardy A, Platt D. Short blocks from the non-coding parts of the human genome have instances within early all known genes and relate to biological processes. PNAS. 2006; 103:6605–6610. [PubMed: 16636294]

Roeder K, Bacanu SA, Wasserman L, Devlin B. Using linkage genome scans to improve the power of association in genome scans. Am J Hum Genet. 2006; 78:243–252. [PubMed: 16400608]

Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D. Calibrating a coalescent simulation of human genome sequence variation. Genome Res. 2005; 15:1576–1583. [PubMed: 16251467]

Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006; 38:209–213. [PubMed: 16415888]

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007; 445:881–885. [PubMed: 17293876]

Thomas D, Xie R, Gebregziabher M. Two-Stage sampling designs for gene association studies. Genet Epidemiol. 2004; 27:401–414. [PubMed: 15543639]

Wang H, Thomas DC, Pe'er I, Stram DO. Optimal two-stage genotyping designs for genome-wide association scans. Genet Epidemiol. 2006; 30:356–368. [PubMed: 16607626]

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]

Wiltshire S, Morris AP, McCarthy MI, Cardon LR. How useful is the fine-scale mapping of complex trait linkage peaks? Genet Epidemiol. 2005; 28:1–10. [PubMed: 15493062]

Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliot KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, Wellcome Trust Case Control Consortium (WTCCC). McCarthy MI, Hattersley AT. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007; 316:1336–1341. Published online 26 April 2007 10.1126/science.1142364. [PubMed: 17463249]

Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. Nat Rev Genet. 2004; 5:89–100. [PubMed: 14735120]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

TABLE I

Descriptive statistics for primary and fine-mapping analyses for a disease allele frequency of 0.2

Genetic model		Single-SNP analysis						Five-SNP haplotype sliding window analysis					
GRR_{hom}	GRR_{het}	*P*(s)	Mean distance	*P*(i|s)	Mean max r^2	*P*(n|s)	Mean max r^2	*P*(s)	Mean distance	*P*(i|s)	Mean max r^2	*P*(d|s)	Mean max r^2
1.0	1.0	0.073	—	0.306	—	0.694	—	0.044	—	0.369	—	0.631	—
1.5	1.5	0.669	47,935	0.612	0.658	0.388	0.688	0.622	47,658	0.767	0.624	0.233	0.556
1.5	1.0	0.095	147,284	0.370	0.153	0.630	0.100	0.067	146,170	0.443	0.238	0.557	0.174
1.5	1.25	0.337	77,260	0.522	0.550	0.478	0.435	0.249	76,618	0.641	0.489	0.359	0.403
2.0	2.0	0.956	32,162	0.705	0.679	0.295	0.819	0.962	29,761	0.850	0.658	0.150	0.760
2.0	1.0	0.149	122,391	0.449	0.347	0.551	0.218	0.111	121,598	0.518	0.347	0.482	0.268
2.0	1.5	0.831	38,558	0.648	0.660	0.352	0.765	0.800	40,581	0.786	0.614	0.214	0.617
3.0	3.0	0.996	27,585	0.702	0.657	0.298	0.880	0.995	25,236	0.891	0.635	0.109	0.815
3.0	1.0	0.276	80,462	0.495	0.534	0.505	0.403	0.216	84,822	0.613	0.492	0.387	0.368
3.0	2.0	0.974	32,115	0.680	0.657	0.320	0.862	0.987	28,600	0.850	0.641	0.150	0.726

P(s) – probability of a significant signal during primary analysis.

Mean distance – between the primary analysis signal and the disease SNP.

P(i|s) – probability of a significant primary analysis signal increasing following fine-mapping.

Mean maximum r^2 – mean r^2 between the primary analysis signal (for a single–SNP or five-SNP haplotype) and the disease SNP.

P(n|s) – probability of the primary analysis signal remaining unchanged (single–SNP analysis) or decreasing (haplotype analyses).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

TABLE II

Magnitude of primary analysis signal increases following fine-mapping for a disease allele frequency of 0.2

Genetic model		Single SNP analysis				Five-SNP haplotype sliding window analysis			
		Size of increase (order of magnitude)				Size of increase (order of magnitude)			
GRR_{hom}	GRR_{het}	Any	≥1	≥2	≥3	Any	≥1	≥2	≥3
1.0	1.0	306	16	1	0	369	40	1	0
1.5	1.5	612	136	57	22	767	237	52	12
1.5	1.0	370	13	2	0	443	51	2	0
1.5	1.25	522	54	8	6	641	115	17	0
2.0	2.0	705	392	269	211	850	493	265	157
2.0	1.0	449	26	4	0	518	67	3	1
2.0	1.5	648	204	100	51	786	303	95	35
3.0	3.0	702	491	411	347	891	661	464	356
3.0	1.0	495	46	5	0	613	113	10	1
3.0	2.0	680	397	325	266	850	520	288	192

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

TABLE III

Probability of a real gene given an increase in significance (any order of magnitude) of an associated single SNP following fine mapping, for a disease allele frequency of 0.2

Genetic model (H_1)		Ten SNPs somewhere in the genome $P(H_1)$=0.00166			One SNP somewhere under a 20-Mb wide linkage peak $P(H_1)$=0.025			One SNP immediately under a linkage peak $P(H_1)$=0.667		
GRR_{hom}	GRR_{het}	$P(H_1\|i,s)$	$P(H_1\|s)$	*PC* (%)	$P(H_1\|I,s)$	$P(H_1\|s)$	*PC* (%)	$P(H_1\|i,s)$	$P(H_1\|s)$	*PC* (%)
1.5	1.5	0.02662	0.01379	93	0.296	0.177	67	0.970	0.944	2.8
1.5	1.0	0.00234	0.00198	18	0.035	0.030	18	0.738	0.704	4.8
1.5	1.25	0.01161	0.00699	66	0.153	0.098	57	0.934	0.894	4.4
2.0	2.0	0.04306	0.01959	120	0.409	0.235	74	0.982	0.960	2.3
2.0	1.0	0.00444	0.00310	43	0.064	0.046	41	0.843	0.789	6.8
2.0	1.5	0.03469	0.01706	103	0.356	0.211	69	0.977	0.954	2.4
3.0	3.0	0.04460	0.02039	119	0.418	0.243	72	0.982	0.962	2.2
3.0	1.0	0.00905	0.00574	58	0.123	0.082	51	0.917	0.874	4.9
3.0	2.0	0.04233	0.01994	112	0.405	0.238	70	0.982	0.961	2.2

$P(H_1)$ – prior probability of a genetic effect.

$P(H_1|i,s)$ – posterior probability of a genetic effect given a significant primary analysis signal increasing in significance following fine-mapping.

$P(H_1|s)$ – posterior probability of a genetic effect given a significant primary analysis signal.

PC (%) – proportional change in posterior probability of a genetic effect, given by $100[P(H_1|i, s)–P(H_1|s)]/P(H_1|s)$.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

TABLE IV

Probability of a real gene given an increase in significance (any order of magnitude) of an associated Five-SNP haplotype following fine-mapping, for a disease allele frequency of 0.2

Genetic model (H_1)		Ten SNPs somewhere in the genome $P(H_1)$=0.00166			One SNP somewhere under a 20Mb wide linkage peak $P(H_1)$=0.025			One SNP immediately under a linkage peak $P(H_1)$=0.667		
GRR_{hom}	GRR_{het}	$P(H_1\|i,s)$	$P(H_1\|s)$	*PC* (%)	$P(H_1\|I,s)$	$P(H_1\|s)$	*PC* (%)	$P(H_1\|i,s)$	$P(H_1\|s)$	*PC* (%)
1.5	1.5	0.03823	0.02190	75	0.379	0.256	48	0.979	0.964	1.6
1.5	1.0	0.00248	0.00242	3	0.037	0.036	3	0.749	0.745	0.6
1.5	1.25	0.01314	0.00889	48	0.170	0.121	40	0.941	0.915	2.8
2.0	2.0	0.06376	0.03346	91	0.512	0.347	47	0.988	0.977	1.2
2.0	1.0	0.00478	0.00399	20	0.069	0.058	19	0.852	0.828	2.9
2.0	1.5	0.04979	0.02799	78	0.446	0.307	45	0.984	0.972	1.3
3.0	3.0	0.06879	0.03458	99	0.532	0.355	50	0.989	0.977	1.2
3.0	1.0	0.01089	0.00770	41	0.145	0.107	36	0.930	0.903	2.9
3.0	2.0	0.06535	0.03432	90	0.518	0.353	47	0.988	0.977	1.1

$P(H_1)$ – prior probability of a genetic effect.

$P(H_1|i,s)$ – posterior probability of a genetic effect given a significant primary analysis signal increasing in significance following fine-mapping.

$P(H_1|s)$ – posterior probability of a genetic effect given a significant primary analysis signal.

PC (%) – proportional change in posterior probability of a genetic effect, given by $100[P(H_1|i, s)–P(H_1|s)]/P(H_1|s)$.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

TABLE V

Probability of a real gene given an increase in significance (at least one order of magnitude) of an associated single SNP following fine mapping, for a disease allele frequency of 0.2

Genetic model (H_1)		Ten SNPs somewhere in the genome $P(H_1)$=0.00166			One SNP somewhere under a 20Mb wide linkage peak $P(H_1)$=0.025		
GRR_{hom}	GRR_{het}	$P(H_1\|i,s)$	$P(H_1\|s)$	*PC* (%)	$P(H_1\|i,s)$	$P(H_1\|s)$	*PC* (%)
1.5	1.5	0.14744	0.01379	969	0.727	0.177	310
1.5	1.0	0.00234	0.00198	18	0.035	0.030	18
1.5	1.25	0.03341	0.00699	378	0.347	0.098	255
2.0	2.0	0.41587	0.01959	2023	0.916	0.235	290
2.0	1.0	0.00730	0.00310	135	0.102	0.046	122
2.0	1.5	0.24350	0.01706	1327	0.832	0.211	295
3.0	3.0	0.48161	0.02039	2262	0.935	0.243	285
3.0	1.0	0.02359	0.00574	311	0.271	0.082	232
3.0	2.0	0.42342	0.01994	2023	0.919	0.238	285

$P(H_1)$ – prior probability of a genetic effect.

$P(H_1|i, s)$ – posterior probability of a genetic effect given a significant primary analysis signal increasing in significance following fine-mapping.

$P(H_1|s)$ – posterior probability of a genetic effect given a significant primary analysis signal.

PC (%) – proportional change in posterior probability of a genetic effect, given by $100[P(H_1|i, s)–P(H_1|s)]/P(H_1|s)$.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

TABLE VI

Probability of a real gene given an increase in significance (any order of magnitude) of an associated single SNP following fine-mapping, for a disease allele frequency of 0.2, stratified by r^2 between the disease SNP and the primary analysis signal

Genetic model		Ten genes in the genome $P(H_1)=0.00166$						One gene under a 20-Mb linkage peak $(H_1)=0.025$					
		$r^2<0.5$		$r^2\geq0.5$		$r^2<0.5$		$r^2<0.5$		$r^2\geq0.5$		$r^2\geq0.8$	
GRR_{hom}	GRR_{het}	$P(H_1\|i,s)$	*PC* (%)	$P(H_1\|i,s)$	*PC* (%)	$P(H_1\|i,s)$	*PC* (%)	$P(H_1\|i,s)$	*PC* (%)	$P(H_1\|i,s)$	*PC* (%)	$P(H_1\|i,s)$	*PC* (%)
1.5	1.5	0.02410	96	0.02793	92	0.02513	74	0.275	71	0.306	66	0.284	54
1.5	1.0	0.00220	13	0.00413	68	0.00345	48	0.033	13	0.060	65	0.051	46
1.5	1.25	0.00929	42	0.01421	90	0.01367	84	0.126	37	0.181	74	0.176	70
2.0	2.0	0.04742	156	0.04140	107	0.03746	86	0.434	93	0.399	67	0.374	56
2.0	1.0	0.00390	27	0.00603	90	0.00546	71	0.057	25	0.085	83	0.078	66
2.0	1.5	0.03584	130	0.03419	93	0.03100	75	0.364	86	0.353	63	0.330	52
3.0	3.0	0.05505	173	0.04075	99	0.03612	76	0.473	96	0.395	62	0.366	50
3.0	1.0	0.00781	38	0.01056	81	0.01001	72	0.108	34	0.141	70	0.135	63
3.0	2.0	0.05278	169	0.03885	94	0.03382	69	0.462	96	0.383	60	0.350	46

$P(H_1)$ – prior probability of a genetic effect.

$P(H_1|s)$ – posterior probability of a genetic effect given a significant primary analysis signal.

PC (%) – proportional change in posterior probability of a genetic effect, given by $100[P(H_1|i, s)–P(H_1|s)]/P(H_1|s)$.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

TABLE VII

Probability of a real gene given no-change in significance of an associated SNP after fine mapping, for a disease allele frequency of 0.2

Genetic model (H_1)		Ten SNPs somewhere in the genome $P(H_1)$=0.00166			One SNP somewhere under a 20Mb wide linkage peak $P(H_1)$=0.025			One SNP immediately under the peak $P(H_1)$=0.667		
GRR_{hom}	GRR_{het}	$P(H_1\|i,s)$	$P(H_1\|s)$	*PC* (%)	$P(H_1\|I,s)$	$P(H_1\|s)$	*PC* (%)	$P(H_1\|i,s)$	$P(H_1\|s)$	*PC* (%)
1.5	1.5	0.00784	0.01379	−43	0.108	0.177	−39	0.905	0.944	−4.1
1.5	1.0	0.00182	0.00198	−8	0.027	0.030	−8	0.686	0.704	−2.6
1.5	1.25	0.00487	0.00699	−30	0.07	0.098	−28	0.855	0.894	−4.4
2.0	2.0	0.00851	0.01959	−57	0.117	0.235	−50	0.912	0.96	−5.0
2.0	1.0	0.00249	0.00310	−20	0.037	0.046	−19	0.75	0.789	−4.9
2.0	1.5	0.00881	0.01706	−48	0.12	0.211	−43	0.914	0.954	−4.2
3.0	3.0	0.00895	0.02039	−56	0.122	0.243	−50	0.916	0.962	−4.8
3.0	1.0	0.00423	0.00574	−26	0.061	0.082	−25	0.836	0.874	−4.3
3.0	2.0	0.00939	0.01994	−53	0.127	0.238	−47	0.919	0.961	−4.3

$P(H1)$ – prior probability of a genetic effect.

$P(H1|s)$ – posterior probability of a genetic effect given a significant primary analysis signal.

PC (%) – proportional change in posterior probability of a genetic effect, given by $100[P(H1|i,s)–P(H1|s)]/P(H1|s)$.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

TABLE VIII

Probability of a real gene given no-change/decrease in significance (any order of magnitude) of an associated 5-SNP haplotype after fine mapping, for a disease allele frequency of 0.2

Genetic model (H_1)		Ten SNPs somewhere in the genome $P(H_1)$=0.00167			One SNP somewhere under a 20Mb wide linkage peak $P(H_1)$=0.025			One SNP immediately under the peak $P(H_1)$=0.667		
GRR_{hom}	GRR_{het}	$P(H_1\|i,s)$	$P(H_1\|s)$	*PC* (%)	$P(H_1\|I,s)$	$P(H_1\|s)$	*PC* (%)	$P(H_1\|i,s)$	$P(H_1\|s)$	*PC* (%)
1.5	1.5	0.00910	0.02190	−58	0.124	0.256	−52	0.917	0.964	−4.9
1.5	1.0	0.00237	0.00242	−2	0.035	0.036	−2	0.741	0.745	−0.5
1.5	1.25	0.00564	0.00889	−37	0.08	0.121	−34	0.872	0.915	−4.7
2.0	2.0	0.00906	0.03346	−73	0.123	0.347	−64	0.917	0.977	−6.1
2.0	1.0	0.00339	0.00399	−15	0.05	0.058	−14	0.803	0.828	−3.0
2.0	1.5	0.01073	0.02799	−62	0.143	0.307	−53	0.929	0.972	−4.4
3.0	3.0	0.00683	0.03458	−80	0.096	0.355	−73	0.892	0.977	−8.7
3.0	1.0	0.00526	0.00770	−32	0.075	0.107	−30	0.864	0.903	−4.3
3.0	2.0	0.00930	0.03432	−73	0.126	0.353	−64	0.919	0.977	−6.0

$P(H_1)$ – prior probability of a genetic effect.

$P(H_1|i,s)$ – posterior probability of a genetic effect given a significant primary analysis signal increasing in significance following fine-mapping.

$P(H_1|s)$ – posterior probability of a genetic effect given a significant primary analysis signal.

PC (%) – proportional change in posterior probability of a genetic effect, given by $100[P(H_1|i,s)-P(H_1|s)]/P(H_1|s)$.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

TABLE IX

Physical localisation of a disease gene during primary single SNP analysis, stratified by r^2 between the disease SNP and the primary analysis signal for a disease allele frequency of 0.2

Genetic model		Mean distance (bp) between disease SNP and primary analysis signal								
		Irrespective of response to fine-mapping			Given an increase (of any size) in significance following fine-mapping			Given no change in significance following fine-mapping		
GRR_{hom}	GRR_{het}	$r^2<0.5$	$r^2\geq0.5$	$r^2\geq0.8$	$r^2<0.5$	$r^2\geq0.5$	$r^2\geq0.8$	$r^2<0.5$	$r^2\geq0.5$	$r^2\geq0.8$
1.5	1.5	100,664	24,796	14,056	74,613	20,445	13,267	143,108	31,563	15,017
1.5	1.0	156,893	48,935	30,560	152,700	44,111	17,498	159,196	54,332	41,918
1.5	1.25	121,886	33,343	18,606	93,768	29,904	20,809	144,483	38,442	15,570
2.0	2.0	56,711	23,582	17,489	43,642	20,622	15,019	118,861	29,394	21,085
2.0	1.0	152,666	36,674	19,957	145,009	31,897	16,653	157,696	43,771	23,794
2.0	1.5	80,326	22,794	15,133	58,303	21,857	15,894	140,964	24,296	14,178
3.0	3.0	45,724	21,113	14,912	37,063	18,408	12,496	112,992	25,846	18,004
3.0	1.0	125,724	26,900	17,760	108,189	23,185	14,269	139,046	31,822	21,885
3.0	2.0	48,561	26,837	17,994	37,679	21,810	14,118	122,973	34,978	22,476