Add intro comment on types of studies leading to the LOD scores or p-values for the GSMA. In general one can have any test stat? — see p.5

Genome Search Meta-Analysis method

The Genome Search Meta-Analysis (GSMA) method [9] was developed to circumvent some common problems of performing meta-analysis on genome-wide linkage studies. The GSMA is a non-parametric method, with few restrictions or assumptions, so that any genome-wide linkage search can be included, regardless of study design or statistical analysis method.

Add a comment regarding association studies
(a) Does GSMA work for these?
(b) Can/should assoc studies be included in a MA with linkage studies? } discuss. ? R ?

In the GSMA, the entire genome is divided into bins of approximately equal cM width. We conventionally use 120 bins of 30cM length, so that for chromosome 1, the region between 0 and 30cM is assigned to bin 1.1, between 30-60cM to bin 1.2, etc. Let the number of bins be n, and the number of studies be m. For each study, the maximum LOD score (or minimum p-value) within each bin is identified, and the bins are ranked, with the most significant result achieving a rank of n, the next highest result a rank of (n -1), etc. Across studies, the ranks for each bin are summed, and the summed rank forms the test statistic for this bin. A high summed rank implies that the bin has high LOD scores within individual studies, and may contain a susceptibility locus. Under the null hypothesis of no linkage, the summed rank for a bin will be the sum of m ranks, randomly chosen from $\{1, 2, \ldots n\}$ with replacement. Significance levels for each bin can be determined from the distribution function of summed ranks [9], or by simulation.

(a) Include sex chromosomes?
(b) Add eg. region 7.3
(c) what to do when chrom doesn't partition evenly into the $l = 30cM$ regions?
(d) no overlap across chrom?

Is there a preference? On what parameters does the sampling distn depend

Under no linkage, the probability of attaining a summed rank R in a specific bin, from m studies and n bins is:

$$P\left(\sum_{i=1}^{m} X_i = R\right) = \begin{cases} 0 & \text{for } R < m \\ \dfrac{1}{n^m}\displaystyle\sum_{k=0}^{d}(-1)^k \binom{R-kn-1}{m-1}\binom{m}{k} & \text{for } m \le R \le mn \\ 0 & \text{for } R > mn \end{cases}$$

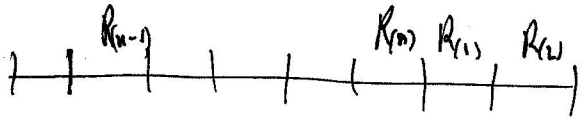
where X_i = rank of study i, d = integer part of $(R-m)/n$ [9]. Hence the probability of obtaining a summed rank of R or greater (i.e. the p-value) in a bin can be calculated. This bin-wise p-value, p_{SR}, can also be obtained by simulation, permuting the bin-location of the assigned ranks. For each study, the ranks within a study are randomly re-assigned to bins, and then the summed rank calculated for each bin. For d replicates, dn summed rank values are obtained, and the p-value for the observed summed rank is calculated from the number of simulated bins with summed rank greater than the observed summed rank (R). The p-value is then $p_{SR} = (r+1)/(dn+1)$, where n is the number of simulated bins[10]. Calculating critical values from simulations is particularly appropriate where the assigned ranks depart from the integer values 1, 2, … n assumed in the distribution function above, through tied ranks or missing values (see Table 1).

The GSMA was developed to encompass diverse study designs and analysis methods. The linkage evidence may be extracted from any analysis method: for example, multipoint LOD scores calculated at each 1 cM, LOD scores calculated at each marker genotyped with the bin, or parametric LOD scores calculated at a series of recombination fractions for each marker. For parametric LOD scores, linkage is often tested using a series of models with different modes of inheritance or different penetrance/frequency parameters. The evidence for linkage can be assessed across all

models analysed, provided the underlying distribution of LOD scores is approximately equal in each model; this can be determined from the distribution of LOD scores across the genome. Thus, the maximum evidence for linkage within a bin would be the highest LOD score calculated, regardless of the model under which it was obtained.

The bin-wise summed rank p-value (p_{SR}) assesses the information in each bin independently of each other bins, and should therefore be corrected for multiple testing. With 120 bins, under no linkage, 6 bins would be expected to attain $p_{SR} < 0.05$, and 1.2 bins to attain $p_{SR} < 0.01$. Following Lander and Kruglyak,[11] we define genome-wide evidence for linkage as that expected to occur by chance once in 20 GSMA studies, and suggestive evidence for linkage as that expected to occur once in a single GSMA study[12]. Using a Bonferroni correction on 120 bins gives p=0.00042 (=0.05/120) for genome-wide significance, and p=0.0083 (=1/120) for suggestive evidence of linkage.

For a genome-wide assessment of linkage, the ordered rank (OR) p-value (p_{OR}) may be used.[12] This uses simulations of the complete GSMA to compare the summed rank of the observed *kth*-highest bin, with the simulated distribution of summed ranks of the *kth* highest bin, i.e. compares the 'place' of the bins in the full listing of results. So, in a simulation of 5000 complete GSMAs, the bin with the highest summed rank is compared to the all 5000 bins with highest summed rank, and the ordered rank p-value (p_{OR}) calculated. Similarly, the summed rank of the bin in the *kth* 'place' is compared to summed ranks of bins lying in *kth* place. This test can identify evidence for many bins with increased evidence for linkage, although the evidence for linkage



within each bin may be modest. In the study of 20 genome wide searches for schizophrenia, 12 bins in the weighted analysis had significant summed rank and significant ordered ranks (p_{SR} <0.05, p_{OR} <0.05). Our simulations based on these studies showed that this combination of significant results was highly unlikely to occur by chance (not observed in 1000 GSMA simulations of an unlinked study). The combination of a significant p_{SR} and p_{OR} is therefore highly predictive of a linkage within a bin, however empiric criteria for linkage for an arbitrary number of studies have not yet been developed.[12] Is there a recommendation for multiple testing correction for ordered p-values?

In assessing linkage we recommend the following hierarchy for interpreting results:

1) A genome-wide significant summed rank p-value (p_{SR} <0.05/#bins)
2) Nominal evidence for linkage in both statistics (p_{SR} <0.05, p_{OR} <0.05)
3) Nominal evidence for linkage in the summed rank (p_{SR}<0.05)

No evidence for linkage should be declared where bins do not have a significant summed rank p-value. Within bins with a significant summed rank, a significant ordered rank p-value can be considered to enhance the evidence for linkage. Clearly, if the *k*th bin has nominal evidence for linkage under both statistics, then any bin with higher summed rank must also be considered significant. For example if three bins attain p-values of 0.011, 0.012, 0.013, the bin with p=0.013 is most likely to have a significant ordered rank, but clearly all bins show significant evidence for linkage. By plotting the observed summed ranks by size, with the distribution of ordered ranks, a 'scree slope' may be seen where the summed ranks decrease rapidly and the ordered ranks become non-significant (see Figure 2, in the inflammatory bowel disease GSMA[13]). In regions where the p_{SR}> 0.05 but p_{OR}<0.05, one interpretation is

and unweighted analysis is performed, using user-defined weights. Three results files are output: (a) results for the most significant bins only, (b) a full genome listing of bin, summed rank, p_{SR}, p_{OR} (weighted and unweighted analyses) (c) ranks assigned to each study, for data checking.

Power to detect linkage using the GSMA

An extensive simulation study of the GSMA was carried out by Levinson et al.,[12] based on genome scans contributed to the meta-analyses of schizophrenia[14] and bipolar disorder[15]. For the simulation, a number of sib pairs with broadly equivalent information to the pedigrees from the original studies were used , with 1625 ASPs for schizophenia, 1017 ASPs for bipolar disorder (narrow phenotype definition), and 501 ASPs for bipolar disorder (very narrow phenotype definition). These three studies therefore give a wide range of study sizes covering those seen in many GSMA studies (Table 3).

The schizophrenia study had high power to detect linkage with a locus conferring a sibling relative risk (λ_s) of 1.3 at a significance level of $p<0.01$. For a significance level of 0.05, a power of at least 70% was attained in the following situations:

- 1625 ASPs (schizophrenia), for a locus with $\lambda_s = 1.15$,
- 1017 ASPs (bipolar disorder, narrow phenotype) for a locus with $\lambda_s = 1.3$,
- 501 ASPs (bipolar disorder, very narrow phenotype) for a locus with $\lambda_s = 1.4$.

Full details of other assumptions required in the simulation, including the number of genotyped parents, marker density, and number of loci simulated are given in the original paper.[12]

This seems to be ill-defined here. In any event, something is unclear.

The power of a study to detect linkage depends on the number of studies (m) and the number of bins (n), in addition to the genetic effect size in each study. The average rank threshold for declaring genome-wide, suggestive or nominal linkage changes with the number of studies (m = 4, 7, 10, 15, 20) and the number of bins (n =60, 120), as shown in Figure 1. Note that the thresholds for genome-wide (p_{GW}) and suggestive (p_{SUG}) linkage depend on the number of bins used: p_{GW}=0.00042 and p_{SUG}=0.0083 for 120 bins, and p_{GW}=0.00056 and p_{SUG} =0.017 for 60 bins; nominal evidence for linkage was fixed at p=0.05 throughout. With 120 bins, an average rank threshold for nominal linkage is 32 for 4 studies, but over 48 for 20 studies – so the average rank is not even within the top third of reported ranks. An average rank of 32 gives nominal evidence for linkage with 4 studies, but provides genome-wide evidence for linkage with 20 studies. With 60 bins, lower average ranks are required for linkage, so that the evidence must be stronger in linked bins where wider bins are used. Provided the maximum LOD scores for a locus localise to a narrow region, using narrow bins provides the most evidence for linkage: with 10 studies, an average rank of 20 gives genome-wide evidence for linkage if this is obtained using 120 bins, but only nominal significance with 60 bins. Reducing the number of bins could, however, increase the power to detect linkage if the LOD scores peaks are too widely spread to be contained in a single bin (for example if the locus lies close to a bin boundary), so that the average ranks decrease using fewer bins.

From where do these come? Fig 1?

What reported ranks?

? meaning between 1-40?

Relative to 120 bins, an analysis smaller requires

The setting does not take account of the assumption that the locus is narrowly defined.

is meant, not what

For a given study size, Relative to 120 bins, an analysis with 60 bins requires smaller average ranks for linkage (Fig. 1). Thus, the evidence must be stronger by pooling smaller correlated bins into wider ones.

One critical issue is the loss of information arising when the GSMA divides the genome into discrete bins. Two simulation studies have compared the power of the GSMA to the power of 'mega analysis', based on genotype data from each study. Demple and Loesgren[19] showed that the power of the GSMA was less than the mega-analysis approaches tested, but they applied the Lander and Kruglyak criteria for genome-wide significance, which is much more stringent than using a Bonferroni multiple testing correction (0.05/#bins). Using this appropriate, less stringent, correction, Levinson et al.[12], showed that the power of the GSMA to detect linkage was actually higher than for the analysis of pooled genotypes.

Extensions of the GSMA

Many different diseases have been studied using the GSMA , but little further methodological development has been carried out. Some authors have proposed minor enhancements to the method. For example in their study of celiac disease, Babron et al.[20] used a summed rank function that was a weighted average of the ranks of a bin and two flanking bins. This extends the potential area in which evidence for linkage can be shown, since high linkage statistics in a flanking bin will be included. However, it will also increase the correlation between summed ranks in adjacent bins. An alternative approach to the problem of maximum LOD scores being attained in adjacent bins in different studies is 'pooled bins' used in the rheumatoid arthritis study.[16] Here, adjacent bins are pooled, and the original analysis of n bins is reanalysed as two analyses of $n/2$ bins each, where bins 1+2, 3+4, … are pooled in the first analysis, and 2+3, 4+5 … are pooled in the second analysis. This analysis would be valuable where a true locus lies close to a bin boundary, and the bin-location of

maximum linkage evidence is inconsistent across studies. However, as Figure 1 shows, reducing the total number of bins reduces the power to detect linkage. ← Has argued both ways: #bins ↑ ⇒ ↑ power #bins ↓ ⇒ ↑ power

In their study of cleft lip/palate, Marazita et al.[21] use a series of overlapping bins from 0-30cM, then 10-40cM, 20-50cM, etc. and assess the maximum evidence for linkage across each possible bin. This should give better localisation information, and may determine whether two linkage peaks exist in one region. However, there are unresolved problems of multiple testing.

Recently, Zintzaras and Ioannidis[22] provided a major extension to the GSMA in developing methods to test for heterogeneity of linkage evidence within a bin. Heterogeneity testing is a standard component of meta-analysis in epidemiological studies, where researchers test for evidence of different effect sizes across studies, but has not previously been implemented in the GSMA. Z&I apply these methods directly to the rank statistics of each study, introducing three, highly correlated heterogeneity statistics. The significance of the statistics are assessed by simulation, randomly reassigning the ranks to bins within each study, and recalculating each heterogeneity statistic. The proportion of simulated bins with Q-statistics above the observed value (for high heterogeneity), or below the observed value (for low heterogeneity) is then tabulated for a p-value. Zintzaras and Ioannidis[22] applied the methods to published ranks in GSMA studies of rheumatoid arthritis[16] and schizophrenia.[14] They identify several bins in each study that show evidence for high heterogeneity (different evidence for linkage across studies) or low heterogeneity (consistent linkage evidence). The authors acknowledge that the distribution of the heterogeneity statistics may depend on the summed rank statistic attained within the

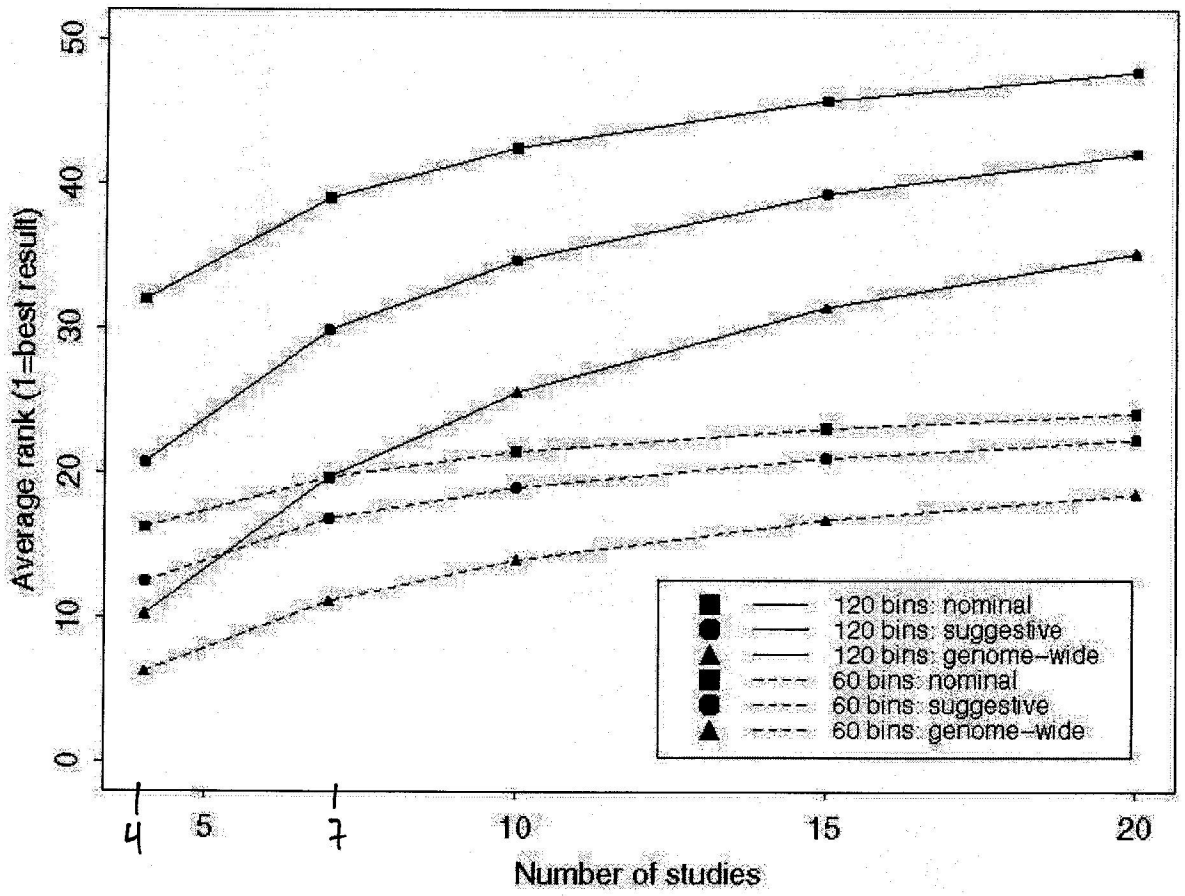
studied. The average number of linkage studies included was 7.9 (range 4-20), and the average number of families was 736 (range 257-1992). (These figures omit the overlapping studies of inflammatory bowel disease, Crohn's disease and ulcerative colitis). Of 14 studies, 8 were full collaborations, while others relied at least partially on published information. All studies found at least one suggestive result (approximately $p<0.01$), and in 12 studies, at least one result of genome-wide significance was found. In the auto-immune diseases, genome-wide significance was found in the HLA region on chromosome 6 (multiple sclerosis,[9] rheumatoid arthritis,[16] psoriasis, [26] inflammatory bowel disease[13]), confirming findings of the original linkage studies. In other studies, a region of genome-wide significance was observed on chromosome 2 for schizophrenia[14], which had not previously been highlighted as a strong candidate region for schizophrenia[28]. Similarly, regions of genome-wide significance were detected on chromosome 4 for psoriasis[26], on chromosome 3 for coronary heart disease[23], on chromosome 2 for cleft lip/palate[21], on chromosome 3 for hypertension[25]and on chromosome 10 for age-related macular degeneration. [17] No susceptibility genes have yet been localised in these regions for these diseases, but they provide strong candidate regions for follow-up linkage or association studies. Genome-wide significance is an extremely stringent criteria (occurring only once in 20 GSMAs by chance), and this is illustrated by the results for Crohn's disease in the region of CARD15 on chromosome 16. This region attained a p-value of 0.003 (weighted analysis)[13], despite the presence of this confirmed susceptibility gene. Across the diseases, there was no correlation between the number of bins with nominal or suggestive significance and the number of studies included. Only five studies had used the Ordered Ranks test to assess clustering of linkage results, but the

Table 3: Summary of published GSMA studies (geno: genotyped individuals, aff: affected individuals, arp: affected relative pairs; asp: affected sibling

List of GSMA studies Disease	First author, year	No. studies	Total no. families	Collaborati	No. of bins	Weighting function	No. bins with summed/average rank Nominal	Suggestive	Genome-wde	pSR<0.05, pOR<0.05	
Multiple sclerosis	Wise, 1999	4	257	No	120	n/d	8	2	1	n/d	
Type 2 diabetes	Demanais, 2003	4	1127	Yes	120	n/d	6	1	0	n/d	
Schizophrenia	Lewis, 2003	20	1208	Yes	120	sqrt(#aff)	12	4	1		12
Bipolar disorder (a)	Segurado, 2003	18	370	Yes	120	sqrt(#aff)	9	2	0		2
Coeliac disease	Babron, 2003	4	442d	Yes	115	#ped	5	5	2	n/d	
Rheumatoid arthritis	Fisher, 2003	4	570	No	120	#asp	10	3	1	n/d	
Coronary heart disease	Chiodini, 2003	4	807	No	124	sqrt(#asp)	4	3	1	n/d	
Inflammatory bowel disease	Williams, 2003	5	709	No	117	n/d	8	4	1	n/d	
Crohn's disease	Williams, 2003	5	472	No	117	n/d	9	4	0	n/d	
Inflammatory bowel disease	van Heel, 2004	10	1253	Yes	105	sqrt(#arp)	8	5	1		6
Crohn's disease	van Heel, 2004	10	711	Yes	105	sqrt(#arp)	10	5	0		8
Ulcerative colitis	van Heel, 2004	7	314	Yes	195	sqrt(#arp)	5	1	0		0
Hypertension/diastolic blood pre	Koivukoski, 2004	9	1992	Yes	120	sqrt#aff)	9	3	1		2
Psoriasis	Sagoo, 2004	6	493	Partial	110	n/d	5	2	2	n/d	
Cleft Lip/Palate	Marazita, 2004	13	574	Partial	120	sqrt(#geno)	*12*	*3*	*1*	12c	
Body mass index	Johnson, 2005	5	505	Yes	121	sqrt(#geno)	n/a	1	0	n/d	
Age-related macular degenerati	Fisher, 2005	6	908	Yes	120	sqrt(#aff)	15	2	1		11

a: very narrow phenotype definition
b: based on fine-scale mapping
c: maximum number, including candidate region follow-up

Average rank threshold for declaring genomewide, suggestive, or nominal linkage by number of studies and number of bins.



To what does this figure correspond? A simulation?

~~Figure legend~~?

Details of simulation given by Levinson et al. (12).