Meta-analysis and Combining Information in Genetics

Rudy Guerra
&
David Allison

Contents

CHAPTER 1

Meta-analysis methods for genome-wide linkage studies

Cathryn M. Lewis
Department of Medical and Molecular Genetics
Guy's, King's and St. Thomas' School of Medicine
King's College London, UK

1.1 Introduction

Genome-wide linkage studies have been extensively used to identify chromosomal regions which may harbour susceptibility genes for complex diseases. The early enthusiasm for such studies has been replaced by the realisation that most complex disease genes have only a minor effect on risk, and consequently many linkage studies have low power to detect such genes (Risch and Merikangas, 1996). This was well illustrated by a compilation of 101 genome-wide linkage studies in 31 diseases, which found that few studies achieved significant evidence for linkage, and there was little replication within each disease (Altmuller et al., 2001). Replication of linkage is an important concept in genome-wide linkage studies: two studies obtaining high (if not significant) LOD scores in the same approximate region lends further weight to these results. This *ad hoc* method of comparing results across studies is formalised in meta-analysis, which provides statistical evidence for the co-localisation of linkage evidence across studies. Meta-analysis can also provide a solution to the lack of power in individual studies: combining weak evidence of linkage from several studies may show an overall significant effect.

Several methods for meta-analysis of linkage studies have been proposed. The gold standard is a complete analysis of genotype data from all contributing studies (often termed 'mega-analysis'). However, many study groups are reluctant to share raw genotype data, particularly if they are restricted by industrial partnerships. There are also technical problems of pooling different marker maps, and difficulties in finding an analysis method that is suitable for all studies. Pooling genotypes in short candidate regions has worked well in many collaborative studies (Demenais et al., 2003; Levinson et al., 2002).

1.2 Statistical methods for meta-analysis of linkage studies

The meta-analysis methods used in epidemiological studies are difficult to apply directly to genetic linkage studies. Methods that pool effect sizes (*e.g.* odds ratios) across studies are inappropriate as linkage studies frequently report results as a test statistic or p-value. In addition, we wish to assess linkage evidence across a region, not at a single location. Novel meta-analysis methods have therefore been developed to take account of the unique design and analysis strategies used in genetic studies.

For a meta-analysis of p-values at a single point, Fisher's method for pooling p-values can be used, provided LOD score values of zero are treated correctly (Province, 2001). However, unless testing for linkage at a strong candidate gene, specifying a single location for the analysis may not be optimal. Simulation studies show that maximum LOD scores have poor localisation, and can arise up to 30cM from a susceptibility gene (Cordell, 2001). Assessing evidence across a region therefore improves the power to detect linkage in a meta-analysis; this strategy is implemented in the Multiple Scan Probability (MSP) method (Badner and Gershon, 2002b). This method extends Fisher's p-value method, using the minimum p-values attained in a region, with a correction to the p-value for the total region length included in the analysis (see below for further details). The meta-analysis of identity-by-descent (IBD) sharing in affected sib pairs has been proposed for both discrete and quantitative traits (Gu et al., 2001) (***see also chapters in this book). Performing meta-analysis on this parameter of effect size is methodologically appealing. However, the IBD sharing statistic is rarely reported in publications, and some methods rely on identical markers being genotyped in each study, which severely restricts their application.

1.3 Genome Search Meta-Analysis method

The Genome Search Meta-Analysis (GSMA) method (Wise et al., 1999) 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.

*** RG: Add intro comment on *types* of studies leading to the lod scores or p-values for the GSMA. In general, can one have *any* test stat?

*** RG: 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)

In the GSMA, the entire genome is divided into bins of approximately equal width (measured in cM). 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.*.

***RG: (a) include sex chromosomes? (b) add *** (can't read) (c) what to do when the chromosome doesn't partition into 30 cM regions? (d) no overlap across chromosomes?

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; 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 (Wise et al., 1999) or by simulation.

***RG: Is there a preference? On what parameters does the sampling distribution depend?

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

***RG: check formula

$$P(\sum_{i=1}^{m} X_i = r) \quad = \quad \begin{cases} 0 & \text{for} \quad r < m \\ \frac{1}{n^m} \sum_{k=0}^{d} (-1)^k \binom{m}{k} \binom{r-kn-1}{m-1} & \text{for} \quad m \leq r \leq mn \\ 0 & \text{for} \quad R > m, \end{cases}$$

where X_i = rank of study i and d = integer part of $(r-m)/n$ (Wise et al., 1999). 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.

***RG: 'bin-location of the assigned ranks' - not quite right wording

For each study, the ranks within a study are randomly re-assigned to bins, and then across studies the summed rank calculated for each bin. For d permutation replicates, dn summed rank values are obtained, and the p-value for an observed summed rank r_{obs} associated with a given bin is calculated from r_{sim}, the number of simulated bins with summed rank greater than or equal to the observed summed rank . The p-value is then $p_{SR} = (r_{sim}+1)/(dn+1)$, where n is the number of simulated bins (North et al., 2003). Calculating critical values by simulation is particularly appropriate when the assigned ranks depart from the integer values $1, 2, \ldots, n$ assumed in the distribution function above, as happens through tied ranks or missing values (see Table 1.1).

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

Table 1.1 *Common sources of incomplete data in the GSMA, and possible solutions*

Missing data problem	Possible solutions
Many bins with a maximum LOD score of zero	Use tied ranks, so 20 bins with a maximum LOD score of zero would be assigned ranks 10.5.
Bins with no genotyped markers or no linkage data	Assign the median rank (*i.e.* $(n+1)/2$ for n bins), or assign a rank which is the weighted average of flanking bins (since multipoint LOD scores are correlated in adjacent bins).
Results are only reported from regions with the strongest evidence for linkage	Contact study authors for full information, and carry out the study collaboratively. Alternatively, if the observed results fall into b bins, assign these ranks $n, n-1, n-2, \ldots, n-(b+1)$, and assign all remaining bins the average remaining rank. For many missing bins, or bins missing in several studies, this method is not advisable, as the distribution function no longer provides a good fit.
Different chromosomes have been included (*e.g.* some studies have not tested the X chromosome)	Analyse all relevant subsets of studies to obtain maximum information, and for each bin/region, report results from the analysis with most complete data. If chromosome X is missing for r studies (out of m), analyse the remaining $m-r$ studies for the whole genome, and report these results from this analysis for chromosome X. Autosomes can then be analysed will all studies.
Two-stage genome wide study, with some regions genotyped on additional families	Use only the first stage analyses: the distribution of the maximum LOD score per bin depends on the number of families included, and a consistent study design should be used across the genome.
High-density genotyping in previously identified candidate regions	Obtain original LOD scores from markers used in the genome search. The maximum evidence for linkage within a bin increases with denser genotyping, thus inflating the evidence for linkage in more densely-genotyped bins.

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 multiple binsand 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 (Lander and Kruglyak, 1995), 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 (Levinson et al., 2003). 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.

***RG: Doesn't seem right; genomewide: 1 in 20 studies, suggestive: 1 in a single study

For a genome-wide assessment of linkage, an ordered rank (OR) p-value (p_{OR}) may be used (Levinson et al., 2003).

***RG: Give some interpretation of ordered p-values?

This uses simulations of the complete GSMA to compare the summed rank of the observed k^{th} highest bin with the simulated distribution of summed ranks of the k^{th} highest bin, *i.e.* compares the 'place' of the bins in the full listing of results. Therefore, in a simulation of 5000 complete GSMAs, the bin with the highest observed summed rank is compared to 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 k^{th} place is compared to summed ranks of all bins lying in k^{th} 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 not consistent with occurring 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 (Levinson et al., 2003).

***RG: Is there a recommendation for multiple testing correction of 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. 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 (van Heel et al., 2004)). In regions where the $p_{SR} > 0.05$ but $p_{OR} < 0.05$, one interpretation is that the power to identify linkage in these bins is low, and a larger meta-analysis might increase significance of p_{SR}, whilst retaining the significance of the ordered rank statistic.

1.4 Collaborative or published information?

Two main approaches are used to carry out a GSMA analysis. Firstly, the GSMA may be based on published information, for example extracting linkage statistics (NPL/MLS scores, p-values, *etc.*) from graphs and tables. In some cases, investigators may have posted detailed genome-wide results or original genotype data on a website. In papers, genome-wide studies are frequently displayed as line graphs of linkage statistics along each chromosome. This may be used in the GSMA by dividing each chromosome into the required number of equal length bins, and reading off the maximum statistic attained in each bin. Inaccuracies in the method arise from different marker maps used in each study, or different chromosome lengths (so that bins will not be exactly compatible across studies). If marker names are given, bins may be designated more accurately by mapping the bin boundary markers relative to the genotyped markers. In some studies, tables of linkage statistics attained at each marker genotyped are given. These markers may be placed into relevant bins, and the maximum linkage statistic for each bin identified. Common problems arising from the use of published data are listed in Table 1.1, with possible solutions.

A more satisfactory method of performing a meta-analysis study is to form a collaboration of relevant research groups, and use computer files of LOD scores (*e.g.* output files generated from Genehunter, Allegro, *etc.*). This gives full information on the location and magnitude of linkage statistic, and should improve the accuracy of the resulting study. However, if some researchers do not wish to participate, the organisers must then choose between an incomplete meta-analysis of high quality data and a complete meta-analysis of lower quality data. In practice, meta-analyses of genetic studies have been widely supported by researchers (*e.g.* schizophrenia (Lewis et al., 2003), bipolar disorder (Segurado et al., 2003), and inflammatory bowel disease (van Heel et al., 2004)).

In any meta-analysis, the investigators rely on the high quality of results generated by the original studies. Any errors due to genotyping problems, inaccurate phenotype definition, incorrect pedigree reconstruction, or poor analysis methods will be carried through to the meta-analysis, and will reduce power to detect evidence for linkage. Errors seem likely to be random in each study, and should therefore not introduce a bias to the meta-analysis results.

1.5 Summed ranks or average ranks?

The GSMA was originally formulated using summed ranks, where the highest rank n is assigned to the bin with the strongest evidence for linkage. This follows the statistical convention that high test statistics (*i.e.* summed rank) show more evidence against the null hypothesis. An alternative, more intuitive, approach is to assign rank 1 to the 'best', most significant bin, and then use the average rank as a test statistic so that low average ranks give stronger evidence for linkage (Levinson et al., 2003). Statistically these approaches are equivalent, and a summed rank of R from n bins and m studies can be converted to an average rank as $(n + 1) - R/m$.

1.6 Bin width

The GSMA is heavily dependent on the chosen bin width. Our original description of the GSMA listed 120 bins, defined by specific boundary markers (see table at `http://www.kcl.ac.uk/depsta/memoge/gsma/` for full marker-bin information). The exact bin width depends on both chromosome length (to give equal width bins on each chromosome) and marker location. Other studies have chosen different bin widths (see Table 1.2). Although narrow bins may intuitively provide more information (see Figure 1.1), localisation through linkage information is broad. Adjacent bins may show evidence for linkage (see, for example, rheumatoid arthritis (Fisher et al., 2003), inflammatory bowel disease (van Heel et al., 2004) GSMA studies) and simulation studies have shown that the strongest information for linkage may arise in the bin flanking the true location (Levinson et al., 2003). In a study of age-related macular degeneration (Fisher et al., 2005), the original 120 bins (of 30cM length) were then bisected, and ranks (for 240 bins) re-assigned to determine whether more bins would improve localisation information or identify novel loci. The results were disappointing, with similar evidence for linkage spreading across several 15cM-width bins, and no novel regions were identified. The relative advantages of narrow or wider bins are listed in Table 1.3.

1.7 Weighted analysis

The original formulation of the GSMA assumed that all studies contributed equally.

However, a study of 500 affected sibling pairs (ASPs) has higher power to detect a true locus than a study of 100 ASPs. This aspect can be reflected in the meta-analysis

Table 1.2 *Summary of published GSMA studies (geno: genotyped individuals; aff: affecteds; arp: affected relative pairs; asp: affected sib pairs; Significance – Nom: nominal; Sugg: suggestive; Gen: genome-wide)*

Disease	Publication	# studies	# families	# bins	Weights	# bins with SR Nom./Sugg./Gen.	$p_{SR} < 0.05$, $p_{OR} < 0.05$
Multiple sclerosis	Wise, 1999	4	257	120	–	8/2/1	–
Type 2 diabetes	*Demanais, 2003	4	1127	120	–	6/1/0	–
Schizophrenia	*Lewis, 2003	20	1208	120	$\sqrt{(\#aff)}$	12/4/1	12
Bipolar disorder[a]	*Segurado, 2003	18	370	120	$\sqrt{(\#aff)}$	9/2/0	2
Coeliac disease	*Babron, 2003	4	442[b]	115	#ped	5/5/2	–
Rheumatoid arthritis	Fisher, 2003	4	570	120	#asp	10/3/1	–
Coronary heart disease	Chiodini, 2003	4	807	124	$\sqrt{(\#asp)}$	4/3/1	–
Inflammatory bowel disease	Williams, 2003	5	709	117	–	8/4/1	–
Crohn's disease	Williams, 2003	5	472	117	–	9/4/0	–
Inflammatory bowel disease	*van Heel, 2004	10	1253	105	$\sqrt{(\#arp)}$	8/5/1	6
Crohn's disease	*van Heel, 2004	10	711	105	$\sqrt{(\#arp)}$	10/5/0	8
Ulcerative colitis	*van Heel, 2004	7	314	195	$\sqrt{(\#arp)}$	5/1/0	0
Hypertension/blood pressure	*Koivukoski, 2004	9	1992	120	$\sqrt{(\#aff)}$	9/3/1	2
Psoriasis	†Sagoo, 2004	6	493	110	–	5/2/2	–
Cleft Lip/Palate	†Marazita, 2004	13	574	120	$\sqrt{(\#geno)}$	12/3/1	12[c]
Body mass index	*Johnson, 2005	5	505	121	$\sqrt{(\#geno)}$	–/1/0	–
Age-related macular degeneration	*Fisher, 2005	6	908	120	$\sqrt{(\#aff)}$	15/2/1	11

* = collaborative study; † = partially collaborative; [a]very narrow phenotype definition; [b]based on fine-scale mapping;
[c]maximum number, including candidate region follow-up

Table 1.3 *Comparison of properties affecting choice of bin width*

Property	Narrower bins (*e.g.* 120 x 30cM bins)	Wider bins (*e.g.* 60 x 60cM bins)
Bin width	Little variability	Unequal bin widths for different length chromosomes
Correlation in ranks in adjacent bins	Highly correlated, particularly for multipoint linkage analysis. May violate distributional assumptions for test statistic.	Low correlation
Localisation	Reasonable, although adjacent bins may be significant	Poor
Power to detect linkage	High, except where maximum LOD scores occur in different bins	Lower, except where wider bins substantially increases the study rank in linked regions
Consistency of bin definition across studies	Poor, especially based on published information	More overlap between bins in adjacent studies, even when poorly defined

by weighting the studies by sample size. The function sqrt(#genotyped affected individuals) has been used in many studies (see Table 1.2) and increased the power to detect linkage by approximately 7% compared to unweighted analyses in a simulation study based broadly on studies in the schizophrenia GSMA (Levinson et al., 2003). The optimal weighting function is unclear, particularly when some studies have used extended pedigrees and others have used ASPs. The power to detect linkage will depend on the locus effects (mutation frequency, penetrance), and for some loci, extended pedigrees may have higher power to detect linkage while affected sib pairs may be the optimal sampling unit for other genes. Defining a single weighting parameter is therefore somewhat unsatisfactory.

The chosen weighting function can be standardised by its average value for all studies, so that the mean weight is 1. Using a narrow range of weights (*e.g.* 0.9 – 1.1) will give an analysis that is very close to the unweighted analysis. However, using one study with a very high weight (*e.g.* four studies with weights 3.0, 0.4, 0.3, 0.3) will give results close to those obtained in this single study. Both these situations should be avoided, and alternative weighting functions may need to be tested.

1.8 GSMA software

Software to perform GSMA on genome-wide linkage studies is available from `http://www.kcl.ac.uk/depsta/memoge/gsma/` (Pardi et al., 2005). This program is written in C++ and available on Windows, Mac, and Unix/Linux platforms. The data input is a table of maximum linkage statistics for each bin, for each study. The program allows for an arbitrary number of bins and studies. Missing values are permitted, and bins replaced with the median linkage statistic for that study. For studies reporting p-values, the entry values should be $1 - p$-value to ensure correct ranking of results. The program calculates the summed rank, then determines the summed rank and ordered rank p-values (p_{SR}, p_{OR}) by simulation. The user may determine the number of simulations, and the program is rapid, completing 10,000 simulations in under 3 seconds on a desktop PC. Weighted 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), and (c) ranks assigned to each study, for data checking.

1.9 Power to detect linkage using the GSMA

An extensive simulation study of the GSMA was carried out by Levinson et al. (2003) based on genome scans contributed to the meta-analyses of schizophrenia (Lewis et al., 2003) and bipolar disorder (Segurado et al., 2003). 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 1.2).

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 $p < 0.01$.

***RG: ‘detect linkage’ - bin containing the disease gene?

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 λ_s = 1.15,
- 1017 ASPs (bipolar disorder, narrow phenotype) for a locus with λ_s = 1.3,
- 501 ASPs (bipolar disorder, very narrow phenotype) for a locus with λ_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 (Levinson et al., 2003).

***RG: (below): ‘power’ seems ill-defined, or at least 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 1.1. 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.

***RG: where do the thresholds come from? Fig 1? What reported ranks?

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.

***RG: meaning between 1 and 40?

An average rank of 32 gives nominal evidence for linkage with 4 studies, but provides genome-wide evidence for linkage with 20 studies. For a given study size, relative to 120 bins an analysis with 60 bins requires smaller average ranks for linkage (Figure 1.1). Thus, the evidence must be stronger by pooling smaller correlated bins into wider ones. 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.

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

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.

***RG: Does the figure correspond to a simulation? (Details of simulation given by Levinson et al).

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 pooling the raw genotype data from each study. Demple and Loesgren (Dempfle and Loesgen, 2004) 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. (2003) showed that the power of the GSMA to detect linkage was actually higher than for the analysis of pooled genotypes.

*** RG/DG: !!! This result seems surprising and possibly counter-intuitive and requires additional comment.

*** RG: Also see Guerra and Goldstein papers

1.10 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.

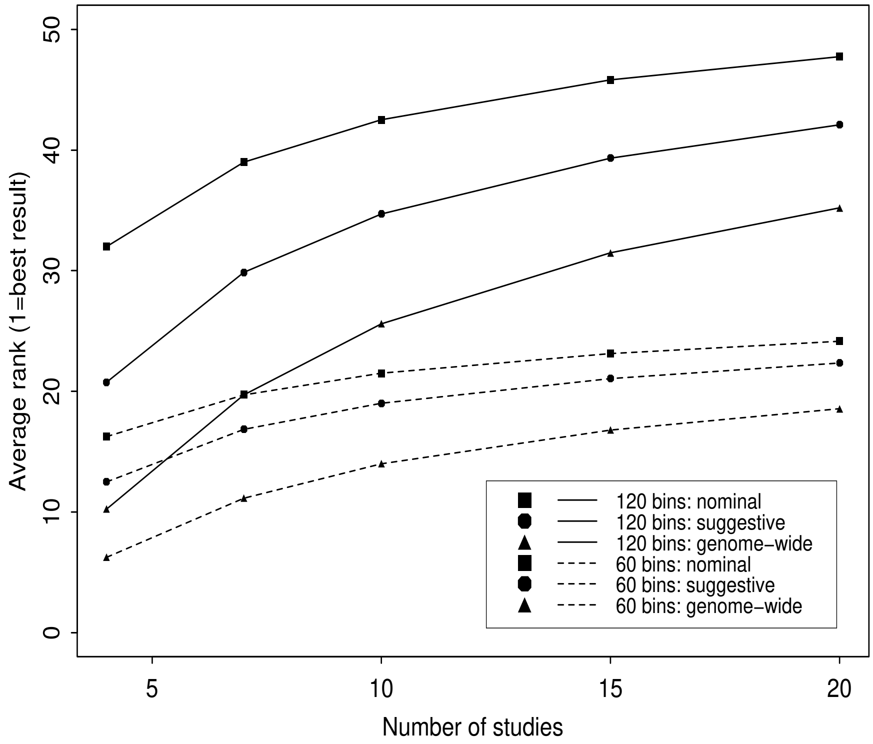

Figure 1.1 *Critical values of the average rank required for genome-wide, suggestive, and nominal evidence for linkage, by number of bins.*

(2003) 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 (Fisher et al., 2003). 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 1.1 shows, reducing the total number of bins reduces the power to detect linkage.

***RG: Has argued both ways: increasing power with increasing number of bins, increasing power with decreasing number of bins.

In their study of cleft lip/palate, Marazita et al. (2004) 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 (2005) 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. They apply these methods directly to the rank statistics of each study, introducing three highly correlated heterogeneity statistics. The significance of each statistic is 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 (2005) applied the methods to published ranks in GSMA studies of rheumatoid arthritis (Fisher et al., 2003) and schizophrenia (Lewis et al., 2003). 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 bin. They therefore test for heterogeneity under two scenarios: where the observed heterogeneity statistic is compared to all simulated bins, and where the observed heterogeneity statistic is only compared to simulated bins with similar summed rank values (± 2).

1.11 Limitations of the GSMA

Three classic sources of error in meta-analysis studies are listed below and discussed with their relevance to the GSMA.

1.11.1 File drawer problem

This error arises when unpublished studies are not included in the meta-analysis, as their existence is unknown to the investigators. For linkage studies of candidate regions, a publication bias exists as negative studies are less likely to be published, which will bias the results of the meta-analysis. For genome-wide studies this is not a major concern: these studies are large, expensive to perform, and publishable, regardless of the significance of LOD scores obtained. No single hypothesis is being tested, so publication bias is not relevant.

1.11.2 Garbage in, garbage out

Any meta-analysis is reliant on the quality of both the data and the results from the individual studies. We assume that each study has a high quality of phenotype and genotype data, and that standard quality control checks have been performed (*e.g.* testing for non-paternity, genotyping errors). The most challenging problem in the GSMA is ensuring a consistent bin definition, particularly where studies have used marker maps that differ in order or distance.

1.11.3 Apples and Oranges

Pooling data from many different studies is statistically appealing, but it is only of value if a common effect is occurring across the studies. There are several sources of heterogeneity that can limit the value of a meta-analysis of genetic linkage studies. Potential sources of heterogeneity are population, family sampling units (extended pedigrees or affected sibling pairs), and clinical characteristics (diagnostic criteria, age of diagnosis, severity of disease). Heterogeneity for evidence of linkage can be tested using the methods of Zintzaras and Ioannidis (2005). A subset analysis can also be performed to analyse a more homogeneous set of studies. We have little understanding of how the distribution of genetic variants contributing to complex disease may be affected by these features, although the common disease, common variant (CDCV) hypothesis for complex diseases implies that a variant would be present across a wide range of study designs. Some GSMA studies have detected linkage to several genetic regions (schizophrenia, inflammatory bowel disease), suggesting that at least some common disease genes can be detected across diverse studies.

1.12 Disease studies using the GSMA

The GSMA has been applied in 14 studies of complex diseases, summarised in Table 1.2 (Demenais et al., 2003; Wise et al., 1999; van Heel et al., 2004; Lewis et al., 2003; Segurado et al., 2003; Fisher et al., 2003, 2005; Babron et al., 2003; Marazita et al., 2004; Chiodini and Lewis, 2003; Williams et al., 2002; Koivukoski et al., 2004; Sagoo et al., 2004; Johnson et al., 2005). Most studies have analysed qualitative diseases, but quantitative traits (hypertension, body mass index) have also been studied. The average number of linkage studies included per meta-analysis 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.

***RG: This p-value adjusted for multiple testing?

In the auto-immune diseases, genome-wide significance was found in the HLA region on chromosome 6 (multiple sclerosis (Wise et al., 1999), rheumatoid arthritis (Fisher et al., 2003), psoriasis (Sagoo et al., 2004), inflammatory bowel disease (van Heel et al., 2004)), confirming findings of the original linkage studies. In other studies, a region of genome-wide significance was observed on chromosome 2 for schizophrenia (Lewis et al., 2003), which had not previously been highlighted as a strong candidate region for schizophrenia (O'Donovan et al., 2003). Similarly, regions of genome-wide significance were detected on chromosome 4 for psoriasis (Sagoo et al., 2004), on chromosome 3 for coronary heart disease (Chiodini and Lewis, 2003), on chromosome 2 for cleft lip/palate (Marazita et al., 2004), on chromosome 3 for hypertension (Koivukoski et al., 2004) and on chromosome 10 for age-related macular degeneration (Fisher et al., 2005). 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) (van Heel et al., 2004), 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 easy availability of this method in the GSMA software package (Pardi et al., 2005) should make this analysis more widely used.

RG: 'clustering'??? (DG: is there a comment here?)

These results show that the GSMA can play an important role in synthesizing data across genome-wide linkage studies and directing follow-up studies. The number of significant regions arising from GSMA studies has raised enthusiasm for the potential utility of linkage studies, these studies suggest that susceptibility genes for complex diseases are detectable using linkage studies, provided the sample sizes are large enough.

1.13 The Multiple Scan Probability method (MSP)

Badner and Gershon (2002b) developed a novel method of meta-analysis of linkage data, based on the maximum evidence for linkage obtained within a genetic region. This method is 'region-wide' rather than genome-wide, as the region for analysis can be specific by investigators, and is usually triggered by one low p-value within a study (*e.g.* $p < 0.01$). For each study, the strongest evidence for linkage within 30cM of the triggering-locus is noted, and the p-values combined, accounting for the length of the region of the final analysis and the genotyping density of original studies (see Badner and Gershon (2002b) for full details). A replication analysis excluding the original linkage finding is also recommended.

This method has been applied to autism (Badner and Gershon, 2002b), schizophrenia and bipolar disorder (Badner and Gershon, 2002a). In schizophrenia, significant evidence for linkage was detected on chromosome 8p, 13q and 22q. These regions on chromosome 8p and 22q were also detected in the GSMA study of schizophrenia (Lewis et al., 2003), but the 13q region was absent. Linkage to 13q and 22q were also found in bipolar disorder, neither of which was detected in the GSMA study (Segurado et al., 2003), however for both schizophrenia and bipolar disorder, the studies included in the GSMA and the MSP differed substantially.

The major contrast between the GSMA and the MSP methods is in the test statistic. The MSP uses a p-value, and therefore retains the magnitude of the significance of the original study. In contrast, the GSMA is a non-parametric rank method, and the maximum contribution from any study is the maximum number of bins (i.e. rank 120 in a study of 120 bins). The MSP should therefore have higher power to detect regions which have strong evidence for linkage in some studies, but with genetic heterogeneity present. Interestingly, the analysis of heterogeneity in the schizophrenia GSMA showed significant genetic heterogeneity on chromosome 13q, which may contribute to the different GSMA and MSP meta-analysis results in this region (Zintzaras and Ioannidis, 2005). The MSP would have lower power to detect regions where linkage evidence is moderate in all studies, as this would not trigger the investigation of a region.

1.14 Conclusions

Millions of dollars have been spent on linkage studies of complex genetic disorders, but the results have been overwhelmingly disappointing. In hindsight, many of these studies are under-powered to detect linkage to genes that confer only a modest increase in risk for a complex disease. However, the utility of linkage studies has been demonstrated by the localisation of a few genes (*e.g.* CARD15 in inflammatory bowel disease, NRG1 in schizophrenia, CAPN10 in type 2 diabetes) following fine-mapping of regions detected in linkage analysis. Linkage studies still have an important role in localising disease genes: genotyping of many large cohorts is in progress, and linkage studies are still widely published. Meta-analysis of linkage studies is therefore a timely approach. It provides a rapid and cost-effective method to ensure that maximum information is extracted from the many linkage studies already performed. The regions highlighted in meta-analysis of linkage can be used to prioritise future gene localisation studies, whether these are based on fine-scale linkage, on association studies of candidate genes, or on follow-up of whole genome association studies.

1.15 Acknowledgements

I thank Douglas Levinson, Sheila Fisher and Lesley Wise for invaluable discussions on the GSMA and meta-analysis in genetics, Fabio Pardi for developing the GSMA software, and all researchers of genome-wide linkage studies for generously contributing their data to meta-analysis studies.

Bibliography

Altmuller, J. et al., Genomewide scans of complex human diseases: True linkage is hard to find, *American Journal of Human Genetics*, 69:936–950, 2001.

Babron, M.C. et al., Meta and pooled analysis of European coeliac disease data, *European Journal of Human Genetics*, 11:828–834, 2003.

Badner, J.A. and E.S. Gershon, Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia, *Molecular Psychiatry*, 7:405–411, 2002a.

Badner, J.A. and E.S. Gershon, Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7, *Molecular Psychiatry*, 7:56–66, 2002b.

Chiodini, B.D. and C.M. Lewis, Meta-analysis of 4 coronary heart disease genome-wide linkage studies confirms a susceptibility locus on chromosome 3q, *Arteriosclerosis Thrombosis and Vascular Biology*, 23:1863–1868, 2003.

Cordell, H.J., Sample size requirements to control for stochastic variation in magnitude and location of allele-sharing linkage statistics in affected sibling pairs, *Annals of Human Genetics*, 65:491–502, 2001.

Demenais, F. et al., A meta-analysis of four european genome screens (GIFT consortium) shows evidence for a novel region on chromosome 17p11.2-q22 linked to type 2 diabetes, *Human Molecular Genetics*, 12:1865–1873, 2003.

Dempfle, A. and S. Loesgen, Meta-analysis of linkage studies for complex diseases: An overview of methods and a simulation study, *Annals of Human Genetics*, 68:69–83, 2004.

Fisher, S.A. et al., Meta-analysis of genome scans of age-related macular degeneration, *Human Molecular Genetics*, 14:2257–2264, 2005.

Fisher, S.A., J.S. Lanchbury, and C.M. Lewis, Meta-analysis of four rheumatoid arthritis genome-wide linkage studies - confirmation of a susceptibility locus on chromosome 16, *Arthritis and Rheumatism*, 48:1200–1206, 2003.

Gu, C., M.A. Province, and D.C. Rao, Meta-analysis of genetic studies, in Rao, D.C. and M.A. Province, eds., *Genetic Dissection of Complex Traits: Challenges for the Next Millennium*, pp. 255–272, San Diego: Academic Press, 2001.

Johnson, L. et al., Meta-analysis of five genome-wide linkage studies for body mass index reveals significant evidence for linkage to chromosome 8p, *International Journal of Obesity*, 29:413–419, 2005.

Koivukoski, L. et al., Meta-analysis of genome-wide scans for hypertension and blood pressure in Caucasians shows evidence of susceptibility regions on chromosomes 2 and 3, *Human Molecular Genetics*, 13:2325–2332, 2004.

Lander, E. and L. Kruglyak, Genetic dissection of complex traits – guidelines for interpreting and reporting linkage results, *Nature Genetics*, 11:241–247, 1995.

Levinson, D.F. et al., No major schizophrenia locus detected on chromosome 1q in a large multicenter sample, *Science*, 296:739–741, 2002.

Levinson, D.F. et al., Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: Methods and power analysis, *American Journal of Human Genetics*, 73:17–33, 2003.

Lewis, C.M. et al., Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia, *American Journal of Human Genetics*, 73:34–48, 2003.

Marazita, M.L. et al., Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32-35, *American Journal of Human Genetics*, 75:161–173, 2004.

North, B.V., D. Curtis, and P.C. Sham, A note on the calculation of empirical P values from Monte Carlo procedures, *American Journal of Human Genetics*, 72:498–499, 2003.

O'Donovan, M.C., N.M. Williams, and M.J. Owen, Recent advances in the genetics of schizophrenia, *Human Molecular Genetics*, 12:R125–R133, 2003.

Pardi, F., D.F. Levinson, and C.M. Lewis, GSMA: software implementation of the genome search meta-analysis method, *Bioinformatics*, 21:4430–4431, 2005.

Province, M.A., The significance of not finding a gene, *American Journal of Human Genetics*, 69:660–663, 2001.

Risch, N. and K. Merikangas, The future of genetic studies of complex human diseases, *Science*, 273:1516–1517, 1996.

Sagoo, G.S. et al., Meta-analysis of genome-wide studies of psoriasis susceptibility reveals linkage to chromosomes 6p21 and 4q28-q31 in Caucasian and Chinese Hans population, *Journal of Investigative Dermatology*, 122:1401–1405, 2004.

Segurado, R. et al., Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder, *American Journal of Human Genetics*, 73:49–62, 2003.

van Heel, D.A. et al., Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs, *Human Molecular Genetics*, 13:763–770, 2004.

Williams, C.N. et al., Using a genome-wide scan and meta-analysis to identify a novel IBD locus and confirm previously identified IBD loci, *Inflammatory Bowel Diseases*, 8:375–381, 2002.

Wise, L.H., J.S. Lanchbury, and C.M. Lewis, Meta-analysis of genome searches, *Annals of Human Genetics*, 63:263–272, 1999.

Zintzaras, E. and J.P.A. Ioannidis, Heterogeneity testing in meta-analysis of genome searches, *Genetic Epidemiology*, 28:123–137, 2005.