Aus dem Institut für Medizinische Biometrie und Epidemiologie
des Fachbereichs Medizin der Philipps-Universität Marburg
Direktor: Prof. Dr. rer. nat. Helmut Schäfer

Evaluation of methods for meta-analysis of genetic linkage studies for complex diseases and application to genome scans for asthma and adult height

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Humanbiologie
(Dr. rer. physiol.)

dem Fachbereich Medizin
der Philipps-Universität Marburg
vorgelegt von

Astrid Dempfle
aus Freising

Marburg 2006

Angenommen vom Fachbereich Medizin der Philipps-Universität Marburg
am 21.12.2006

Gedruckt mit Genehmigung des Fachbereichs

Dekan: Prof. Dr. Bernhard Maisch
Referent: Prof. Dr. Helmut Schäfer
Koreferent: Prof. Dr. Karl-Heinz Grzeschik
2. Koreferent: Prof. Dr. Johannes Hebebrand, Universität Duisburg-Essen

Table of contents

1 Introduction

Why Meta-Analysis?

In recent years, the focus of genetic epidemiology has been on diseases that are common and thus of major interest to public health. These are mostly complex diseases that have several genetic and environmental influences. The contribution of each disease gene to the overall risk is presumed to be small, thus large sample sizes are necessary to identify such loci by linkage analysis. The power of samples that can reasonably be recruited within a few years in one center is therefore limited. The popular affected sib pair (ASP) design with a few hundred ASPs is known to have low power for genes with small to moderate effects (e.g. Risch and Merikangas 1996) and many examples exist where genome scans did not show significant evidence for linkage (Altmüller et al. 2001). However, for many complex diseases several groups worldwide recruited families and the initial analyses of many genome scans are already published, so meta-analysis seems to be a realistic scenario now. A combined sample of similar studies could result in a sample size sufficient to find loci with smaller effects. Therefore researchers might be willing to cooperate towards a joint analysis.

Differences between studies

Different studies usually vary in many aspects, and the heterogeneity caused by numerous sources should be taken into account. These are, for example, different ascertainment schemes and family structures (population based family cohorts, nuclear families with affected sib pairs, extended pedigrees with many affecteds or even one large pedigree from a population isolate). Typically another obvious difference is the sample size. Whereas pooling automatically accounts for sample size differences between studies, this is not the case for methods that combine p-values or effect estimates. For genotyping markers, different “standard panels” are used, e.g. from the Marshfield Mammalian Genotyping Center (see

http://research.marshfieldclinic.org/genetics) or commercial marker sets (such as the ABI Prism linkage mapping sets, see http://www.appliedbiosystems.com), sometimes with additional markers in candidate regions or genes. These panels vary in number and spacing of markers and their heterozygosity. In many studies, environmental and other covariables are collected, but not in a consistent manner across studies. These covariates may or may not be included in the initial analysis, but are rarely published in detail. Even the phenotype definition varies between studies, e.g. it may be possible to use either a quantitative phenotype or a discrete trait for the same disease (like hypertension). Especially in psychiatric disorders, the diagnosis is sometimes difficult because no objective measure of disease exists, and different diagnostic schemes are used (such as ICD 10 and DSM IV). The different ethnic background of study populations can lead to genetic heterogeneity between samples, especially in combination with different ascertainment criteria and phenotype definitions. As for a single linkage study, locus heterogeneity seriously diminishes the power of meta-analysis. However, very little is known about its extent in complex diseases, and one would expect that at least some common disease genes can be identified in a combined sample. Hardly any work has been done on the question if and when it is appropriate to aggregate evidence across genetic linkage studies and to test whether samples are homogenous enough. This applies to both genetic homogeneity (same disease gene responsible in all study populations, same phenotype) and variability caused by study design, for instance family structures. A test for homogeneity was proposed by Babron et al. (2001), but its usefulness for the typical sample sizes, where it might have very low power, was questioned by the authors. The differences in study design also pose methodological problems as they preclude simple methods of combining evidence for linkage across studies. For example, just adding model-based two-point LOD scores over families (from different studies) is not possible if different markers at different locations were genotyped, or different genetic models were used for analysis. And traditional methods of fixed or random effects meta-analysis that produce an overall effect size estimate require that the same parameter or effect size measure is estimated and available from all studies, which is usually not the case if different study designs were used.

For genetic linkage studies, a combination of tests of the same null hypothesis is usually required for meta-analysis, namely that there is no linkage at a specific

locus. This implies that genotype data of the same marker must be available in all studies or that equivalent information about the inheritance at the specific locus is available in all studies through multipoint linkage methods (e.g. Kruglyak et al. 1996). But even with multipoint methods, the estimated inheritance pattern depends on the number, positions, and heterozygosity of the genotyped markers in a region (and the availability of genotyped founders), all of which determine the information content. For regions with low information content, the estimated sharing of alleles identical by descent (IBD) between relatives is biased towards the null hypothesis under several commonly used IBD estimation methods (Schork and Greenwood 2004). With different marker panels genotyped, there is a systematic difference in IBD estimates between studies even with the same study design and underlying genetic model, which is neither considered by most published meta-analysis methods, nor by simple pooling. Therefore, pooling of raw data should not be considered the gold standard of meta-analysis (for biases from simple pooling see also Bravata and Olkin 2001).

Previous applications of meta-analyses to linkage studies of complex diseases

Interest in meta-analysis and the awareness of the necessity of collaboration and meta-analysis seems high in the human genetics community (Conneally 2003), and the number of published meta-analyses for complex diseases has been rising steadily over the last years. Examples for meta-analyses of linkage genome scans include a meta-analysis of published genome scan data for autism (Badner and Gershon 2002b), meta-analyses of genome scans for hypertension and blood pressure (Koivukoski et al. 2004), cleft lip/palate (Marazita et al. 2004), psoriasis (Sagoo et al. 2004), rheumatoid arthritis (Choi et al. 2006), body mass index (Johnson et al. 2005) and age-related macular degeneration (Fisher et al. 2005). A two-stage approach has been employed by Demenais et al. (2003) who conducted a meta-analysis of four European genome scans for type 2 diabetes, first using the GSMA method (Wise et al. 1999, which is presented in detail in chapter 2.3.1) for the whole genome as a “screening tool” for the most promising regions and then pooled the individual genotype and phenotype data and family information for a combined analysis of only those chromosomes that were significant in the GSMA. Such a combined analysis of individual data has also been perfomed for whole genome scans on diseases such as multiple sclerosis (The Transatlantic Multiple

Sclerosis Genetics Cooperative 2001) and type I diabetes (Cox et al. 2001) where three groups each combined their samples. It has also been used for a combined analysis of data from eight linkage genome scans for body mass index (Wu et al. 2002). For some diseases, consortia were formed to facilitate and prospectively plan the pooling of data. They have proposed standardized study protocols and consistent methods for diagnosis and data collection (CSGA asthma consortium (Xu et al. 2001), The ADHD Molecular Genetics Network (2000)).

Publication bias

Publication bias, the usual concern about meta-analyses that negative results are less likely to be published, could be a smaller problem in the context of genome scans. These are always large, expensive projects which are bound to be published, and some interesting results are obtained in virtually every scan. Exceptions could be those studies conducted by industry. One special form of publication bias could occur in genome scans when detailed results are only given for regions that show some evidence for linkage and genome wide results are only presented as small figures. Including only those scans or regions with a significant result in a meta-analysis would lead to serious bias. In this case, it is necessary to request the complete and exact results from the authors. Also linkage studies for candidate regions might not be published if negative. Thus, a thorough search for all relevant studies, including unpublished ones or those presented only at scientific meetings is an important first step in a meta-analysis. Tests for publication bias and sensitivity analysis are also advisable (for some general remarks and methods see Sutton et al. 2000).

Objectives

In this thesis, the following objectives will be addressed:

A) Give an overview of existing meta-analysis methods both in general and specifically for linkage genome scans (chapter 2).

B) Describe in detail a new statistical method for meta-analysis of genetic linkage studies (chapter 3). This method was first proposed by Loesgen et al. (2001).

C) Compare the power of the proposed new method with existing methods and explore effects of different markers sets and sample sizes on the power of meta-analysis (chapter 4). This simulation study was published in Dempfle and Loesgen (2004).

D) As a first application of the proposed weighted meta-analysis, a meta-analysis of a binary phenotype, i.e. asthma affection status, is performed (chapter 5). This was carried out in the context of the Genetic Analysis Workshop 12 (GAW 12) and published in Loesgen et al. (2001).

E) A meta-analysis of the quantitative phenotype adult height is performed (chapter 6), using a different meta-analysis methodology. This was motivated by the results of a linkage genome scan of this phenotype in the Framingham Heart Study, which was published as Geller/Dempfle et al. (2003, joint first authorship).

In chapter 7, the implications of this research for meta-analysis of linkage studies and the phenotypes asthma and adult height are discussed.

2 Statistical Methods for Meta-Analysis

2.1 Overview of statistical methods for meta-analysis in general

In the classical meta-analysis context (e.g. clinical trials) there is typically one pre-specified effect of interest for which estimates or tests from several studies are combined (Hedges and Olkin 1985). In genome scans for linkage on the other hand, statistical tests, for example on sharing of alleles identical by descent (IBD) between relatives, are considered at many loci across the whole genome, and the interest is in the location of the largest test statistic. Some methods for the classic case of one specific effect of interest are reviewed first, before coming to the genome scan situation.

The aim of a meta-analysis is to take advantage of an increased sample size for a more precise estimate or for a more powerful statistical test (or both) than in the individual samples. In general, it can be distinguished between p-value based methods, which are purely tests of significance and those to combine effect estimates, often together with the estimation of a confidence interval and a significance test.

Methods based on the combination of p-values

Overviews of p-value methods are given by Folks (1984) and Hedges and Olkin (1985). If k independent studies test equivalent null hypotheses H_{0i}, i=1, ..., k, a combined test of the omnibus null hypothesis H_0: "all H_{0i} are true" can be constructed. Let p_i be the one-sided p-values from continuous test statistics (possibly different statistics in the different studies) used to test the individual H_{0i}. Then the p_i are distributed as independent uniform variables on the interval [0,1] under the null hypothesis. This fact is used for most combined tests. The first such test was proposed by Tippett (1931, as cited by Hedges and Olkin (1985)), which uses the smallest p-value $p_{(1)}$. Under the null hypothesis, the probability that $p_{(1)}$ is

smaller than $1 - (1 - \alpha)^{1/k}$ is α, which gives a level α test. This procedure is admissible for tests from the exponential family (Folks 1984).

Another often used procedure was proposed by Fisher (1932), who transformed the uniformly distributed p-values to χ^2 variables and combine these. If p has a uniform distribution, then -2 ln(p) has a χ^2 distribution with 2 degrees of freedom. So the combined test uses $-2\sum_{i=1}^{k} \ln p_i$, which is distributed as a χ^2 random variable with 2k degrees of freedom. This test procedure is also admissible for tests from the exponential family (Folks 1984). Elston (1991) shows that Fisher's method is asymptotically optimal among essentially all methods of combining independent tests.

Two related modifications of Fisher's method are given by Olkin and Saner (2001) and Zaykin et al. (2002) who propose trimming some p-values, either all above a certain threshold or the n largest (or smallest, with n<k) and give the correct distributions of the resulting test statistics. This method is more powerful than Fisher's original method for alternatives where only some of the k null hypotheses are false, while Fisher's method which uses all p-values will rather detect departures from the omnibus null hypothesis when most of the individual null hypotheses are false.

A third method that uses the uniform distribution of p-values is to transform them via the standard normal distribution (attributed to Stouffer et al. 1949 by Hedges and Olkin (1985)). If $Z_i=\Phi^{-1}(p_i)$ (with Φ^{-1} the inverse of the standard normal distribution function), then $Z = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} Z_i$ is distributed as a standard normal. This can also be used to give studies different weights w_i, as

$$Z = \sum_{i=1}^{k} \frac{w_i Z_i}{\sqrt{\sum_{i=1}^{k} w_i^2}}$$

is also distributed as a standard normal (the unweighted inverse normal method is the case of setting all weights equal to 1). These weights can reflect study characteristics such as sample size with more weight for larger, more precise studies and less weight for small studies. This procedure is also admissible for tests from the exponential family (Hedges and Olkin 1985). Since it also uses all

p-values, it is more powerful against alternatives where most of the individual null hypotheses are false than in cases where only some of them are false.

Methods based on effect estimates

Meta-analysis methods that use effect estimates usually result in a combined effect estimate together with a confidence interval and a significance test. In general, the effect estimates from single studies are weighted proportional to their precision for a combined estimate. Depending on the research question and study designs of the individual studies, appropriate effect size measures, e.g. odds ratios or relative risks, differences in survival between treatment and control group or the number of alleles shared IBD between sib pairs have to be extracted from each study. Following Sutton et al. (2000), let T_i be the observed effect sizes in i=1, … , k independent studies, each with variance v_i. Let θ_i be the underlying population effect sizes for which T_i are estimates from each study, i.e.

$T_i = \theta_i + e_i$, with e_i being a random sampling error that has mean 0 and variance v_i.

In a fixed effects model, all population effect sizes are assumed equal, i.e.

$\theta_1 = \ldots = \theta_k = \theta$, where θ is the true common underlying effect size. Then a fixed effects meta-analysis estimate of the common effect is

$$T = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i}, \text{ with } \quad w_i = \frac{1}{v_i}.$$

These weights (inversely proportional to the variance in each study) minimize the variance of T, which is then

$$\text{var}(T) = \frac{1}{\sum_{i=1}^{k} w_i}.$$

If a sufficient number of studies are combined, T can be assumed to be normally distributed, which leads to a canonical confidence interval and significance test.

If the assumption of a common underlying effect size θ is not justified, random effects meta-analysis can be used. Here, heterogeneity between studies is included in the model and the study-specific (true) effects θ_i are assumed to come from a random distribution of effect sizes with a fixed mean θ and variance σ_τ^2:

$\theta_i = \theta + \tau_i$.

Therefore the model includes both within and between study variation, with independent error terms:

$$T_i = \theta + \tau_i + e_i .$$

Again, a weighted estimate of θ can be constructed as above now with weights

$$\tilde{w}_i = \frac{1/(\sigma_\tau^2 + v_i)}{\sum_{i=1}^{k} 1/(\sigma_\tau^2 + v_i)}.$$

In practice, estimates of v_i and σ_τ^2 are used for weighting.

Estimation of between-study heterogeneity is an important step in meta-analysis, which is also performed to decide whether studies are combined using fixed or random effects meta-analysis. This is usually done by Cochran's Q-statistic (Cochran 1937 as cited by Sutton et al. (2000)), which is defined as follows using the above notation:

$$Q = \sum_{i=1}^{k} w_i (T_i - T)^2 .$$

Under the null hypothesis of homogeneity, Q is distributed as a χ^2 variable with k-1 degrees of freedom. However, if only a small number of studies is combined such heterogeneity tests usually have very low power (Sutton et al. 2000). Stratified analyses for subgroups of studies defined by plausible causes of heterogeneity may be a useful addition to formal heterogeneity tests.

The results of a meta-analysis are often presented graphically. The most typical form of display is the so called forest plot in which much information can be conveyed succinctly (Sutton et al. 2000). The point estimates of the relevant effect size measure together with a 95% confidence interval are shown for each study, together with the combined estimate and its confidence interval. The size of each study is often represented additionally as the size of the box denoting the point estimate.

2.2 The special situation of meta-analysis of linkage genome scans

In genetic linkage studies, the choice and definition of appropriate effect sizes can be difficult. For affected sib pairs (ASP) the number of alleles shared IBD at a marker locus is often the parameter of interest, e.g. in the popular mean test of Blackwelder and Elston (1985) and related test statistics. However, if the study design includes also other affected relative pairs, e.g. cousins, the underlying effect size measure will be different. For quantitative traits, the effect size can be the Haseman-Elston regression coefficient (Haseman and Elston 1972) which

relates the phenotypic differences between siblings to the IBD sharing. But since there is no general agreement as to which study design should be used for linkage analysis of complex diseases, and even less which test is preferred for a particular study design, the researcher is confronted with the problem that published studies of the same disease use different statistics which are based on completely different effect size measures. Therefore it is often not possible to extract estimates of a common effect size measure from different studies. Besides, the results are usually reported as statistical tests (often termed LOD score, but not necessarily following the same distribution as a parametric LOD score (Morton 1955; Nyholt 2000) and it is often not possible to transform this back to a parameter estimate and its variance. Lastly, since a genome scan includes not only one statistical test but several hundreds along the genome, it is common to report only results that reached some level of significance in detail in the paper and not results for all positions tested. So the necessary information for a meta-analysis usually will have to be requested from the authors of the original studies.

The goal of a meta-analysis of genome scans is not only to get a more powerful test for any specific locus but to identify the loci or genetic regions with the strongest overall evidence for linkage in the genome. Even though technically hypothesis tests are considered, a linkage genome scan is rather a method to generate a hypothesis (determine a region for subsequent analysis, including finemapping and functional studies) than to test one or more specific hypotheses (like the involvement of a particular candidate gene). The results of genome wide analyses are never taken as a definite proof for the involvement of a specific gene. Therefore, claiming linkage to a region which in the end turns out to be false is often considered less serious than missing a true linkage.

The growing interest in meta-analysis of linkage studies has led to the adaptation of standard statistical methods and the development of new specific methodology, which will be briefly reviewed below. In general, methods for linkage genome scans can be divided into those for pooling raw data and those for genuine meta-analysis that combine statistics from different studies (see table 1 for some differences between pooling and meta-analysis). Meta-analysis methods can be further divided into those that combine significance tests (e.g. in the form of p-values) and those that combine effect estimates and then test the significance of the estimated common effect.

Table 1: Some differences between pooling and meta-analysis of genetic linkage studies.

Pooling	Meta-Analysis
▪ Needs complete raw data (family structures, genotypes and phenotypes of all individuals)	▪ Needs only intermediate results, e.g. LOD scores (but these usually for all loci)
▪ All available information can potentially be used	▪ Only aggregated information is available
▪ Automatically accounts for sample size	▪ Possible to account for sample size
▪ Often only possible for same markers	▪ Methods to account for different markers can be developed
▪ Can use same linkage analysis method	▪ Has to cope with different linkage methods
▪ Can use methods that account for family structures, covariates or heterogeneity on an individual or family level	▪ Can use such information only on a per-study basis, not individually

Pooling of linkage genome scan data would mean to use the complete family structures and genotypes and phenotypes of all individuals in all studies in one combined analysis which treats data as if it was derived from one single study, while meta-analysis uses only the results, such as LOD scores or effect estimates along the genome, of each study. Pooling of raw data automatically uses all available information and thus might be the most powerful way to combine linkage data, but this is not necessarily the case. Multipoint methods will enable the combined analysis even if different marker sets were genotyped in the different studies. However, without consideration of systematic differences, pooling may yield spurious or paradoxical results (Bravata and Olkin 2001). Especially in unbalanced designs, this can lead to counterintuitive conclusions, if e.g. in clinical trials a treatment is beneficial in two (unbalanced) studies, and a pooling of the data shows that in the pooled study groups the treatment does not appear beneficial. This phenomenon is known as Simpson's paradox (Simpson 1951, as cited by Bravata and Olkin 2001) and can be avoided by weighting results appropriately before combining. In linkage analysis, unbalanced designs may not be such a serious problem, but systematic genetic heterogeneity between studies may be a concern that is better addressed by meta-analysis than simple pooling. But even for a meta-analysis, the availability of the complete genotype data (instead of just LOD scores or other summary data) would be helpful, since this enables a potentially sensible primary analysis with the same linkage statistic

and facilitates incorporation of study characteristics (such as family structures) or covariates into the meta-analysis.

2.3 Overview of existing methods for meta-analysis of linkage genome scans

2.3.1 Meta-analysis methods specifically for linkage studies

2.3.1.1 Genome Search Meta-Analysis (GSMA)

A meta-analysis method specifically developed for linkage genome scans is the Genome Search Meta-Analysis (GSMA) by Wise et al. (1999), which is a non-parametric rank method to evaluate the combined evidence for linkage in several genome scans. As input from each study, it allows any linkage test statistic and, for each chromosomal region, measures the relative significance of the results across different studies. For GSMA, the chromosomes are divided into bins of roughly equal size, each bin containing at least one measurement per study. Thus the complete results of the genome scan have to be available, not just the locations and LOD scores of significant linkage regions. The authors suggest approximately 30 cM intervals as bins to combine typical human genome scan results, yielding about 120 bins. For each scan, the result corresponding to the smallest level of significance in each bin is recorded; this could be the highest LOD score in the interval or the smallest p-value. Within each study, the bins are ranked according to this result, and the ranks within a bin are summed across all studies. Wise et al. (1999) give the distribution of this rank sum under the null hypothesis that there is no susceptibility locus within the bin. This assumes that results of different bins are stochastically independent. No combined effect estimate can be computed with this method. The different studies do not need to be analyzed with the same statistical method (e.g. the results of parametric LOD score analyses can be combined with non-parametric allele sharing statistics), and it is not necessary to have the same markers genotyped in different studies. Wise et al. (2001) extend their method to work with studies not covering the whole genome, e.g. candidate gene studies, but in this situation the distribution under the null has to be obtained by simulation. The same would be necessary to incorporate weights, e.g. for study size, which is not considered in the originally proposed method. As the procedure is very computer intensive for a simulation study, the investigation in the following chapter is restricted to unweighted GSMA.

2.3.1.2 Multiple scan probability (MSP)

Badner and Goldin (1999) proposed a method to combine p-values across studies by calculating the probability for a set of attained p-values in a genetic region. The procedure is proposed for situations where not the complete results of the whole genome scans are available, but just the usually published smallest p-values with their locations. Each p-value is corrected for multiple testing by accounting for the length of the investigated interval using the dependence structure of genetic linkage along a chromosome (Lander and Schork 1994). So if for one study the smallest p-value attained on a certain chromsome arm is known, this will be corrected for the genetic length of this chromosome arm. The resulting adjusted p-values are then combined using Fisher's procedure. As such, this type of meta-analysis does not take into account different sample sizes between studies. This method cannot yield a location estimate, but will give an adjusted meta-analysis p-value which applies to the largest region included in the meta-analysis. This method was further refined and applied to bipolar disorder and schizophrenia (Badner and Gershon 2002a) and also to autism (Badner and Gershon 2002b).

2.3.2 Methodological adaptations and applications of general meta-analysis methods to linkage studies

2.3.2.1 Fisher's method for combining p-values

For the evaluation of several independent tests of the same null hypothesis, the before mentioned procedure of combining p-values by Fisher (1932) is available. Under the null hypothesis, that there is no disease gene at a particular locus, p-values p_1, ... , p_n from n independent studies are uniformly distributed on the interval [0,1], and -2 ln p_i is distributed as a χ^2 random variable with 2 degrees of freedom. Therefore, the combination of p-values, $-2\sum_{i=1}^{n} \ln p_i$, is distributed as a χ^2 random variable with 2n degrees of freedom. When using this method, studies can be analyzed with different statistical methods, i.e. p-values can be derived from different test statistics, but have to be available at the same loci.

A note on the use of Fisher's method in the context of linkage analysis was given by Province (2001). Since many nonparametric linkage methods truncate the LOD score space at 0 (i.e. do not give evidence *against* linkage), he suggested

interpreting a nonparametric LOD of 0 as a p-value of 0.72 (instead of 0.5), to avoid bias when including results from such studies.

Fisher's method to combine p-values (Fisher 1932), has been applied in genetic epidemiology by Allison & Heo (1998) who used this method to combine p-values from single point analyses at different markers across a candidate region in studies of obesity. Guerra et al. (1999) compared Fisher's method with the pooling of raw data on Genetic Analysis Workshop 11 simulated data (Greenberg et al. 1999). Finally, Wu et al. (2002) analyzed eight samples from four ethnic groups in a study on obesity and used Province's modification (2001) of Fisher's method (interpreting a nonparametric LOD of 0 as a p-value of 0.72). A similar approach was applied recently in a meta-analysis of genome scans for fasting glucose, insulin, and insulin resistance (An et al. 2005).

2.3.2.2 Truncated product method (TPM)

Zaykin et al. (2002) present a generalization of Fisher's method for combining p-values in the context of genetic analyses. For this method, only p-values below a certain threshold τ (e.g. 0.05, naturally accounting for habits of reporting only 'significant' results) are considered. They derive the exact distribution of the product of these p-values under the null by conditioning on the number of p-values less than τ. This is derived from the fact that the number of p-values less than τ (from k independent tests) follows a binomial distribution Bin(k, τ) under the global null hypothesis. Again, these p-values can be obtained by different methods but must be given for each study on every marker intended for meta-analysis. TPM has been proposed for genetic studies, but has not been applied to linkage genome scans yet.

2.3.2.3 Fixed and random effects meta-analysis

The general concept of obtaining combined estimates of relevant effect size measures through fixed or random effects meta-analysis have also been proposed in the context of genetic linkage analysis, both for single locations (e.g. candidate genes) and whole genome scans. They require the availability of sensible effect estimates and their variances for each study at each genetic location. These are generally not part of the published data for genome scans and can rarely be

extracted from the published data (as is usually the case for other types of studies, e.g. clinical trials). Therefore such methods are only feasable if the complete individual data for each study are available to conduct sensible primary analyses which yield such parameter estimates for all samples. As mentioned above, even if the raw genotype data from several studies is available, it might be sensible to use a meta-analysis which accounts for possible heterogeneity instead of pooling the data in one combined analysis without regard for systematic differences such as different genotyped marker sets. Approaches that aim at combining effect estimates in model-free linkage analysis mostly use the number of alleles identical by descent (IBD) between relative pairs as the common effect across studies (e.g. Wu et al. 2002). Gu et al. (1998) present a method how to derive and combine IBD estimates even from different sib pair designs (e.g. concordant affected and discordant) in a random effects model. Further developments of this approach are found in Gu et al. (1999) and Gu et al. (2001). Similarly, McQueen et al. (2005) used the complete genotype data of eleven linkage studies of Bipolar Disorder to perform the same statistical analysis for each sample, test heterogeneity of IBD estimates and finally combine them in a random-effects model. Goldstein et al. (1999) proposed combining IBD estimates in a fixed effects meta-analysis, by weighting estimates with the inverse of their variance. For quantitative phenotypes, a sensible effect size measure from linkage analysis can be the Haseman-Elston regression coefficient (Haseman and Elston 1972). These were combined in fixed and random effects meta-analyses by Etzel and Costello (2001), Iyengar et al. (2001) and Jacobs et al. (2001) in the context of the Genetic Analysis Workshop 12 (Meyers et al. 2001) where the complete, individual data of four genome scans and five candidate studies on asthma were available. Etzel and Guerra (2002) investigated this approach in some more detail. A different effect size measure for quantitative traits was used by Hejimans et al. (2005), who initially analyzed serum lipid levels in each of four samples separately using an inverse regression method (Sham et al. 2002; implemented in Merlin-Regress, Abecasis et al. 2002), which yields locus-specific heritabilities and corresponding standard errors. These were then combined across samples by use of a random effects meta-analysis, but without further testing of heterogeneity.

2.3.3 Heterogeneity tests for linkage studies

In a traditional meta-analysis framework, estimation of between-study heterogeneity is an important step. This is usually done by Cochran's Q-statistic (Cochran 1937 as cited by Sutton et al. (2000)) which was introduced in chapter 2.1, or variants of this. The specific form of this heterogeneity test if the effect of interest is the mean IBD sharing among affected sib pairs has been given e.g. by Gu et al. (1998; 2001) and Goldstein et al. (1999).

This type of heterogeneity test needs the study specific effect estimates and variances, just as the estimation of a common effect in a fixed or random effects meta-analysis. Therefore it is usually only possible to perform for linkage genome scans if the original genotype data are available from all studies. If no effect size estimates but just measures of significance (e.g. LOD scores) are available from each genome scan, such a test is not possible. In theory, it is possible to use the fact that a one-tailed p-value is a function of sample size and effect size, to construct a heterogeneity test using just p-values, but this is only valid if all studies have the same sample size (Hedges and Olkin 1985, p 126) and thus of limited use. Similar approaches to inferring an effect estimate from a p-value have been proposed (Rosenthal and Rubin 2003) but these are subject to analogous inadequacies (Kraemer 2005; Hsu 2005). So in practice a heterogeneity test based just on measures of significance is not adequate if different study designs, test statistics and sample sizes are used and would be of limited applicability.

In a similar spirit, Zintzaras and Ioannidis (2005) proposed a generalization of Cochrans Q statistic for the GSMA method, using ranks instead of effect estimates in each of 120 bins. Ranks are only measures of the relative significance of a genomic region in relation to all other regions in the same study. As such, ranks for each bin across studies are independent of sample size, study design or statistical test used. This test is supposed to be performed for each bin, leading to 120 heterogeneity tests for a genome-wide meta-analysis. For one bin, let R_i and w_i be the rank and weight for study i and R the average rank across all studies, then the test statistic would be

$$Q = \sum_{i=1}^{k} w_i (R_i - R)^2.$$

Two similar metrics were also proposed, and in all cases, the significance of the test statistic was assessed by permutations. The power of this approach (e.g. in

comparison to using effect size measures has not been assessed. This procedure for heterogeneity testing has been used by Malhotra et al. (2005) in a meta-analysis of genome scans for lipid traits in African Americans.

An additional concern in genome-wide linkage studies is the question for which loci heterogeneity tests should be performed, since studies usually differ in the markers that are genotyped. One approach would be to test heterogeneity for each locus for which a combined linkage test (or effect size estimate) is intended, e.g. at every cM position along the genome, thus leading to approximately 3600 tests for a whole genome scan (based on sex-averaged genetic length of the human genome (Kong et al. 2002)). But since such homogeneity tests usually have very low power if only a small number of studies is combined (Sutton et al. 2000), performing so many tests might lead to an unacceptably high amount of both false positive and false negative results. Other options have also been proposed, such as one test per chromosome arm to account for the considerable variation in location of the maximum linkage peak in small samples. Therefore just the maximum linkage statistic at each chromosome arm for each study could be included in a test for heterogeneity (Babron et al. 2001).

In conclusion, heterogeneity tests for linkage genome scans have low power, and are rarely feasible as they need the complete individual data from all studies. Additionally, not enough research has been done on the question for which loci such tests should be best performed. Stratified analyses, e.g. based on subgroups of studies of the same ethnicity or population may be more appropriate.

2.3.4 Graphical display of results

A graphical display of meta-analysis results in relation to individual results can be done in a similar fashion as the presentation of the primary genome scan results, using a graph of the obtained LOD scores or p-values against the genetic location along the genome (for examples see the applications in chapters 5 and 6). This makes it possible to simultaneously depict the results for the complete genome. However, contrary to the graphical presentation of meta-analysis results in traditional cases where just one effect is of interest, the precision of the individual studies will not be apparent from such a display. Results for the most interesting loci (such as those reaching significance) could be presented in such a way, e.g. as

a forest plot, if the relevant effect size estimates and confidence intervals are available for all studies.

3 Proposed new method: weighted combination of Z-scores

For genetic linkage analysis of complex traits, non-parametric methods are often used, which do not require specification of a genetic model. These model-free methods evaluate the number of alleles shared IBD between affected relatives. One of the most popular study designs for linkage analysis of complex diseases is the affected sib pair (ASP) design, where families with at least two children affected with the disease of interest are ascertained. The children and their parents (if available) are genotyped with highly polymorphic microsatellite markers or large numbers of single nucleotide polymorphism (SNP) markers spread throughout the genome. Non-parametric linkage analysis is then performed, e.g. using the mean test of Blackwelder and Elston (1985), which is based on the number of alleles shared IBD between an ASP. Such tests are implemented in commonly used software such as Genehunter (Kruglyak et al. 1996), Allegro (Gudbjartsson et al. 2000) and Merlin (Abecasis et al. 2002). The popular NPL_{pairs} statistic as implemented in Genehunter is a generalization of the mean test, which is equivalent to the mean test for ASPs and a fully informative marker and generalizes to other family structures such as larger sibships or affected cousins. For a marker which is not fully informative in a given family, the genotypes at neighboring markers are used to estimate IBD sharing, based on the known dependence structure of linkage along a chromosome. Such multipoint methods, using Hidden Markov Models are used to incorporate information from flanking markers and allow the calculation of the linkage statistic also for positions between markers (Kruglyak et al. 1995).

In an ASP study, let $\pi_i(t)$ be the number of alleles shared IBD by sib pair i at location t (in cM). Then $\pi_i(t)$ follows a trinomial distribution under the null hypothesis of no linkage and takes values 0, 1 and 2 with probabilities ¼, ½ and ¼. Thus the expectation of $\pi_i(t)$ is 1 and its variance ½. For most genetic locations $\pi_i(t)$ is not directly observable, so likelihood-based methods were developed to

estimate $\pi_i(t)$ conditional on all marker data on one chromosome (Risch 1990b; Kruglyak et al. 1996). The NPL_{pairs} test statistic at location t for a study j (j = 1, ..., k) with n_j independent ASPs is then defined as $Z_{pairs,\,j}(t)$:

$$Z_{pairs,\,j}(t) = \frac{1}{\sqrt{n_j}} \sum_{i=1}^{n_j} \frac{E(\pi_i(t)|marker\ data)\text{-}1}{\sqrt{1/2}} \ .$$

Under the null hypothesis of no linkage at locus t, this is the standardized mean allele sharing. Thus, $Z_{pairs,\,j}(t)$ is asymptotically distributed as a standard normal (Kruglyak et al. 1996), if the inheritance pattern at this position can be completely inferred (fully informative marker).

For meta-analysis using effect estimates, $\pi(t)$, the mean number of alleles shared IBD between ASPs is the appropriate effect size for which a combined estimate is required. $Z_{pairs,\,j}(t)$ is a standardized estimate of this variable for study j, multiplied by $\sqrt{n_j}$. The variance of $\pi(t)$ is proportional to $1/n_j$, so weighting this estimate with n (or equivalently $Z_{pairs,\,j}$ with $\sqrt{n_j}$) is optimal to minimize the variance of the combined estimate as in a fixed effects meta-analysis (Hedges and Olkin 1985). Thus, a combined linkage test statistic can be defined, by weighting each study specific NPL statistic with $\sqrt{n_j}$.

At a marker that is not fully informative, the expectation under the null hypothesis is used to estimate the IBD sharing in Z_{pairs}, which makes it a conservative test and leads to a biased estimate of IBD sharing under the alternative (Kong and Cox 1997; Schork and Greenwood 2004). In the extreme case, where a family is completely uninformative (e.g. not genotyped for a marker) this leads to an estimated IBD sharing of 1 for an affected sib pair, as would be expected under the null hypothesis. This so called “perfect data approximation” (Kruglyak et al. 1996) will always result in reduced power for locations between genetic markers, since here the inheritance pattern has to be estimated. For single studies, this is also evident in the fact that NPL scores always peak at markers (see figure 3 in chapter 4) and drop between markers, contrary to parametric LOD scores which are usually higher between markers. Since NPL scores are preferentially evaluated at the genotyped markers in individual studies, this power loss between markers is most relevant if marker spacing is large (and heterozygosity low). For meta-analysis however, studies

which used different sets of markers are included and thus results at all genetic locations are relevant. Uncertainty on the inheritance pattern therefore reduces power for non-parametric linkage analysis and can be interpreted as reducing the effective sample size of a study (Guo and Elston 2000). Examples for this loss of power caused by ambiguity of the inheritance pattern at a genetic location are described in the simulation study and shown in figures 2 and 3 (chapter 4). Unweighted pooling of NPL scores across studies, or weighting just with study size, is therefore not optimal when different markers are genotyped and the amount of information on inheritance in families varies between studies at each locus. A measure for the uncertainty of the inheritance pattern at a genetic location is the information content (IC) as first proposed by Kruglyak and Lander (1995) and refinded by Kruglyak et al. (1996). This uses the information-theoretic entropy as a measure of residual uncertainty in a probability distribution, here the distribution of inheritance vectors for a pedigree, which describe the inheritance pattern at a genetic location within a family. The entropy at a genetic location *t* is defined as

$$E(t) = -\sum_i P_i(t) \log_2 P_i(t)$$

where $P_i(t)$ is the probability of the i^{th} possible inheritance vector for a family. In the absence of genotypic data, the probability distribution is uniform over all 2^{2n-f} equivalence classes of inheritance vectors (with n denoting the number of non-founders and f the number of founders in a family), thus the entropy in this case is $E_0=2n-f$. The information content for a family is then defined as

$$IC(t) = 1 - E(t)/E_0 \ .$$

The entropy is an additive measure and is summed over all pedigrees in a sample. With this definition, IC is a general measure of available information on the inheritance pattern in the included families which does not depend on any specific linkage test statistic and always lies between 0 and 1, taking the value 1 for a fully informative marker and the value 0 in the absence of any genotype information (Kruglyak et al. 1996). The product of IC and sample size therefore is a reasonable approach to approximate the effective sample size. Using IC to approximate the effective sample size has the advantage that this is standard output of most multipoint linkage analysis programs (e.g. Genehunter, Allegro or Merlin) and therefore may be available for a meta-analysis, while other measures

of information which are based on individual genotypes will not be accessible for meta-analysis.

Recently, the loss of power in non-parametric linkage analysis caused by less than perfect information about inheritance has also been discussed in detail by Schork and Greenwood (2004) which lead to a lively debate (Cordell 2004; Mukhopadhyay et al. 2004; Visscher and Wray 2004; Sieberts et al. 2004; Abecasis et al. 2004). As several authors pointed out, the extent of this power loss depends on the test statistic used and on the specific implementation in a software algorithm. Specifically, the NPL_{pairs} statistic (as implemented in Genehunter (Kruglyak et al. 1996)) and the equivalent mean test (as implemented e.g. in SIBPAL/S.A.G.E. 2004) suffer from a relevant loss of power due to the "perfect data approximation" used that treats uninformative (or partly informative) families inappropriately. This is overcome by non-parametric linkage statistics which were proposed by Kong and Cox (1997), termed allele sharing LOD scores, which are implemented e.g. in Genehunter-Plus (Kong and Cox 1997), Allegro (Gudbjartsson et al. 2000) and Merlin (Abecasis et al. 2002). Franke and Ziegler (2005) suggested a weighting scheme for individual families to improve the mean test, by weighting each affected sib pair with the appropriate marker informativity. A comparison of the power of their proposed approach with the power of the Kong and Cox (1997) non-parametric LOD scores has not been done, yet. A comparison of this test statistic with Holmans' possible triangle test (Holmans 1993) and other extensions of the mean test (Knapp 2006) shows that it performs poorly for realistic genetic models, especially in situations with low marker informativity, for which it was designed (Knapp 2006). Sensible linkage test statistics therefore should not treat un-informative or partly informative families as though they would provide evidence for the null hypothesis of no linkage but rather consider that they provide only limited or no information at all for the test, thus reducing the effective sample size (Mukhopadhyay et al. 2006). If such test statistics are used, the loss in power is reduced. However, the effective sample size of a study is still diminished if families in the sample are un-informative or partly informative for a marker. And even if genetic markers are highly polymorphic and all individuals within a family are genotyped, the multipoint approximation used to calculate the test statistic even for locations between

markers (which have to be considered in meta-analysis), always leads to a reduction of the effective sample size.

Consequently, a combination of NPL scores is proposed which uses the following different weights: study size (equals pooling of the raw data in the case of ASPs), information content, and if the information content is not available an exponential function of the distance to the next marker (which provides a simple means to describe the loss in information content between markers). These methods for weighted meta-analysis were first proposed in Loesgen et al. (2001) and further refined and evaluated in Dempfle and Loesgen (2004).

For any locus t let $IC_j(t)$ denote the information content in study j, and d_{tj} be the distance (in cM) from t to the closest genotyped marker in study j. Then the different weighting schemes are as follows:

Unweighted: $Z_{unweighted}(t) = \frac{1}{\sqrt{k}} \sum_{j=1}^{k} Z_{pairs,j}(t)$

Study size: $Z_{size}(t) = \sum_j \frac{\sqrt{n_j} Z_{pairs,j}(t)}{\sqrt{\sum_j n_j}}$

Information content and study size: $Z_{IC+Size}(t) = \sum_j \frac{IC_j(t)\sqrt{n_j} Z_{pairs,j}(t)}{\sqrt{\sum_j IC(t)_j^2 n_j}}$

Distance and study size: $Z_{Dist+Size}(t) = \sum_j \frac{\exp(-0.1 d_{tj})\sqrt{n_j} Z_{pairs,j}(t)}{\sqrt{\sum_j \exp(-0.1 d_{tj})^2 \; n_j}}$

The unweighted combined tests statistic is used for comparison in the simulation study in the next chapter and analogously Z_{IC} and Z_{Dist} are defined without the weighting factor for study size. With these definitions, the weighted scores are again distributed as standard normal under the null hypothesis, so appropriate p-values can be calculated. The normality assumption is confirmed in the following simulation study (chapter 4) on unlinked chromosomes both for weighted and unweighted Z-scores.

The implementation of the proposed meta-analysis is very simple: it uses the NPL scores at regular intervals (e.g. every 1 cM) for each study, which are standard output of all programs for multipoint non-parametric linkage analysis. If the NPL scores are only available at genotyped markers, they may be interpolated

for the desired regular locations between markers. Depending on the desired weighting scheme and extend of available data, it is also necessary to obtain the sample size and study specific IC at the same intervals as the NPL scores (which is also standard output from e.g. Genehunter, Allegro or Merlin). If IC is not accessible, the locations of genotyped markers can be used to calculate the distance between each evaluated position and the closest marker. The combined NPL statistic is then calculated from these ingredients (usually at regular intervals again) and the p-value can be derived from the standard normal distribution.

If other test statistics instead of NPL scores have been calculated in the original studies, these may be included through the inverse normal method. P-values or LOD scores from other test statistics are converted to Z-scores via the normal distribution and can then also be usefully weighted and combined. The application of the more sophisticated weighting schemes with IC or a distance measure is of course dependent on the availability of the necessary data and if these cannot be obtained, a weighting just with sample size will still be possible and valuable.

General concerns regarding tests for heterogeneity in genome-wide linkage studies have been addressed in the preceding chapter and apply to this method as well. Especially low power if only relatively few studies are combined (which will be the case for most diseases) in combination with severe multiple testing make such tests very unreliable (Babron et al. 2001). The parameter of interest in non-parametric ASP studies is the mean IBD sharing and as a test of heterogeneity, Cochran's Q statistic can be used (Gu et al. 1998). However, a valid estimate of the variance of the mean IBD sharing without the assumption that the null hypothesis is true requires the locus-specific, family-wise IBD estimates (McQueen et al. 2006), which are usually not available for a meta-analysis. From just the NPL scores themselves, which are standardized for each study, no valid estimate and variance of the relevant parameter can be calculated, especially because information content and thus effective sample size varies across genetic locations. Another drawback is that in practical applications the sample often contains a mixture of similar but not identical family structures, such as more than two affected siblings or additionally some affected cousins. More distant relatives have lower expected IBD sharing and thus there is not one single relevant parameter (such as $\pi(t)$) that applies to all families or all samples in a meta-

analysis. All this makes a formal heterogeneity test practically infeasible if only summary results (such as NPL scores) and not individual data are available for a meta-analysis. If potential sources of heterogeneity, such as different ethnicities between study samples, are known, a useful strategy would instead be to perform meta-analyses for presumably more homogeneous subsets of samples from the same population as well as for all study samples. Other known study-specific covariates, such as average disease severity of ascertained cases or average age-of-onset, could also be used to define clinically relevant subgroups which may be genetically more homogeneous and perform subgroup specific meta-analyses.

As a graphical display, a plot of the individual and the meta-analysis NPL scores in one figure (for each chromosome or genome-wide) will allow useful comparisons. Examples of such figures can be seen in the applications to the phenotypes asthma and height (chapters 5 and 6). Forest plots, as are used for meta-analyses of just one specific parameter of interest, will not be practical for genome-wide analysis.

The proposed methods are widely applicable for meta-analysis of linkage studies. Multipoint NPL scores are calculated in many affected sib pair studies for complex disease. Other multipoint test statistics can be used via the inverse normal method. The necessity to have results for all loci is not unique to the proposed method, but arises in similar form in any other meta-analysis method for genome wide linkage studies as statistical tests of the same hypothesis, i.e. linkage to the same locus, have to be combined. Here, Badner's method (Badner and Gershon 2002b) is an exception that can be applied to the usually published data which still is often only the significant p-values. GSMA needs at least one value of each study in each bin which is rarely the case if only interesting values are reported. However, more and more researchers distribute the detailed results of their analysis via the internet, like NPL scores for each marker tested (as in a psoriasis scan by Nair et al. 1997), where the multipoint statistics at regular intervals (as in a stature and BMI genome scan by Perola et al. 2001) and the information content could be included. Besides, it is a long standing scientific tradition to keep data and intermediate results and to give other researchers the opportunity to verify the reported results by making materials and data available. Many scientific journals (e.g. Nature, Science, The American Journal of Human Genetics, Human Molecular Genetics, or Molecular Psychiatry) stress this

obligation in their instructions for authors. The need to provide all original data to reviewers and, after publication, to all interested scientists was emphasized in the context of microarray gene expression studies by Perou (2001) and deposition of microarray data in public databases is now explicitly demanded by most journals (MIAME standard, Brazma et al. 2001). His reasoning, that other scientists need the chance to check the alleged conclusions and to gain additional insight by a re-analysis applies equally to genome wide linkage analysis. It should therefore be a matter of course to obtain at least the detailed results for all markers from the authors of a published genome scan.

4 Simulation Study: Exploring Effects of Different Markers and Sample Sizes

4.1 Simulated model

Whole genome scan data was simulated to compare the power of different methods of meta-analysis. The simulations are intended to reflect a realistic scenario, which represents studies by four separate research groups undertaken in the same or closely related populations. This means that a clinically homogeneous sample with the same underlying genetic model of disease, such as disease allele frequencies and penetrances, is simulated for all studies. The simulations correspond to a situation where collaboration between groups is anticipated or planned in advance, so that study protocols, especially with regard to ascertainment and diagnosis, are similar or even standardized. Therefore the affected sib pair (ASP) design was chosen for all studies. The studies differ in their sample size (i.e. the number of ASP) and in the marker panels that are genotyped. The emphasis of this simulation study is on the effect of using different marker panels and the resulting variation in information content at each locus.

A binary trait dependent on a bi-allelic disease locus is modeled. The population frequency of disease allele D is 0.1 and that of the low-risk allele or group of alleles d equals 0.9. The penetrances are 0.4 for genotype DD, 0.1 for genotype Dd and 0.025 for genotype dd, so the disease alleles act multiplicatively with a genotypic relative risk (GRR) of 4. This results in a λ_{sib} of 1.54 and a $\lambda_{offspring}$ of 1.48 (Risch 1990a) and an expected IBD sharing at the disease locus of 0.597 (see table 2). To investigate whether the results generalize to other trait models, two other scenarios were also simulated. These were a recessive model with a very similar effect (expected IBD sharing 0.594, penetrances 0.05, 0.05 and 0.6 for the homozygous wildtype, heterozygous and homozygous disease allele

carriers) and a multiplicative model with a relative risk of only 3.5 (expected IBD sharing 0.576, penetrances 0.05, 0.175 and 0.613).

Table 2: Parameters of different simulation scenarios.

Scenario	Multiplicative GRR 4	Recessive	Multiplicative GRR 3.5
Penetrance DD	0.4	0.6	0.613
Penetrance Dd	0.1	0.05	0.175
Penetrance dd	0.025	0.05	0.05
Prevalence of disease	0.042	0.056	0.078
λ_{sib}	1.537	1.287	1.392
$\lambda_{offspring}$	1.479	1.088	1.360
E(IBD|ASP)	0.597	0.594	0.576

1000 replicates of affected sib pair data of 80 families on four different marker panels A, B, C, D were simulated, which represent the four different studies. Marker loci are spaced every 5 or 15 cM on two 90 cM chromosomes with 2 or 6 equifrequent alleles at each marker (such as SNP or microsatellite markers). See table 3 for the combination of these parameters. The disease gene is always located at position 46 cM of chromosome 1, but as the position of the first marker varies, so does the distance of the closest marker to the disease locus as shown in figure 1.

Table 3: Parameters for the simulation of study types A to D.

Type	Number of families	Distance between markers (cM)	Number of equifrequent alleles	First marker (position in cM)	Distance of closest marker to the disease locus (cM)
A	80	15	6	8	7
B	80	5	2	3	2
C	40+40	15	6	0	1
D	40+40	5	2	0	1

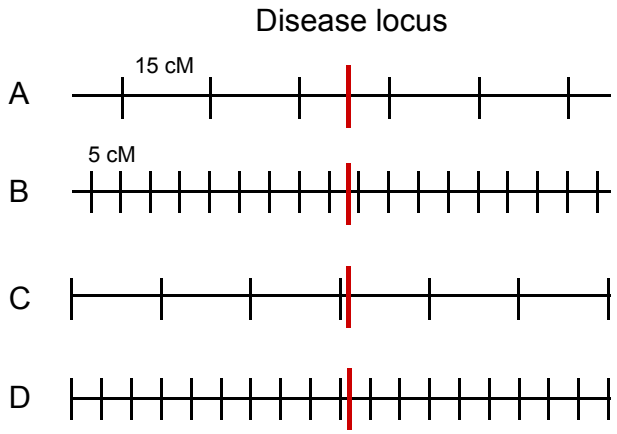


Figure 1: Simulated Marker Maps. Total length is 90 cM, disease locus is at position 46 cM.

Chromosome 2 is "unlinked", i.e. it does not have a disease gene. To investigate the influence of study size, data on marker panels C and D were split into two groups of 40 sib pairs each, referred to as C_1, C_2, D_1, and D_2, respectively. Simulations were performed with the genome scan simulation program SIMSCAN developed by S. Loesgen (unpublished). The core algorithm renders this program especially suited for the evaluation of methodology for affected relative designs. The basis of the following comparisons in the simulation study is a 'standard analysis' of each simulated study with Genehunter NPL_{pairs} (Kruglyak et al. 1996) in 1 cM steps (for ASPs and a fully informative marker, the NPL_{pairs} is equivalent to the mean test of Blackwelder and Elston (1985)). Combined analyses of the four simulated studies were done with the proposed weighting methods for Z-scores (unweighted, weighted by size, information content, distance or combinations of these), with the GSMA, Fisher's method and TPM as described in the previous chapter. As the GSMA method uses only the relative significance (i.e. the rank) of a region compared to all other regions in the respective study, the p-values depend on the number of bins. To get a good approximation of a full genome scan, the results of the linked chromosome in each study are used together with the results of 39 replicates of an "unlinked" chromosome, for a total of 120 bins in each study. As expected this gives indeed higher power than adding just 2 or 10 null chromosomes.

4.2 Results of the simulation study

4.2.1 Effect of recombination and low informativity on power

The power of the different meta-analysis methods to detect linkage to the simulated disease locus is compared. The power of a statistical test under a specific alternative (here the simulated disease model) is determined by the sample size (in this case the number of informative sib pairs). The maximal power of an affected sib pair linkage study with a given sample size, and a given test statistic, will be achieved when the inheritance pattern at the disease locus can be correctly inferred in every family. This is the case if each sib pair is fully informative. In reality, this upper bound will not be reached, as not every meiosis at every marker is informative, and additionally there is recombination between the disease locus and the closest marker. However, this overly optimistic approach is frequently used in sample size calculations (e.g. Risch and Merikangas 1996) as this is a straightforward method and corrections for realistic situations are difficult to model. The theoretical power of the NPL_{pairs} test with full information at the disease locus (which is the maximal power achievable with this test statistic) is given in table 4 for different significance levels and sample sizes for scenario 1 and in table 5 for different sample sizes in scenarios 1 to 3 at a significance level of 0.00074. Under the assumptions mentioned above, a single study of 80 ASP has little power to establish linkage in scenario 1 (less than 25% at $\alpha = 0.00074$, the threshold for "suggestive" linkage as defined by Lander and Kruglyak (1995)), but a combined sample of 320 ASP can reach a good power of about 80% even at a stringent significance level of $\alpha = 0.000022$ ("significant" linkage) and over 95% power at $\alpha = 0.00074$. Scenarios 1 and 2 with different genetic models have very similar expected power, whereas the multiplicative genetic model with the smaller effect in scenario 3 has much smaller power.

Table 4: Theoretical power (in %) of the mean test for ASPs in a single study with the simulated disease parameters of scenario 1 at different significance levels (with a fully informative marker locus completely linked to the disease locus).

Number of ASPs	$\alpha = 0.01$	$\alpha = 0.00074$	$\alpha = 0.000022$
40	27.13	6.98	0.82
80	54.82	22.74	4.73
320	99.55	95.95	79.41

Table 5: Theoretical power (in %) of the mean test for ASPs in a single study of scenarios 1 to 3 at a significance level of 0.0074 (with a fully informative marker locus completely linked to the disease locus).

Number of ASPs	Scenario 1: multiplicative, GRR = 4	Scenario 2: recessive, GRR = 4	Scenario 3: multiplicative, GRR = 3.5
80	22.74	21.05	10.31
320	95.95	94.82	75.44

Recombination between the disease locus and a marker will diminish the chance to detect linkage. The incorporation of recombination rates in the power calculation for the mean test is for example implemented in the TDT-Power Calculator (primarily designed for TDT power calculation, but also capable of calculating the power of the mean test for ASPs) by Chen and Deng (2001). Note that even though marker allele frequencies are required to be specified in this program, they are only used in the TDT power calculation, but not in the power calculation for the mean test, which assumes that the number of alleles identical by descent at the marker locus can be unequivocally inferred for each sib pair, as with a fully informative marker. Table 6 gives the estimated power under scenario 1 with different numbers of ASPs and recombination fractions. A distance of 1 cM between the disease locus and a fully informative marker does not affect the power very much, but a distance of 7 cM will result in an important loss of power, e.g. from about 51% to 30% with 80 ASPs at a significance level of 0.01 or from 73% to 31% for 320 ASPs at a level of 0.000022.

Table 6: Power (in %) of the mean test for ASPs in a single study under scenario 1 at different recombination fractions between disease locus and a fully informative marker (calculated with the TDT-Power Calculator) at different significance levels.

Number of ASPs	Distance (cM)	$\alpha = 0.01$	$\alpha = 0.00074$	$\alpha = 0.000022$
80	1	50.90	19.91	3.86
80	7	30.04	8.31	1.07
320	1	99.21	93.89	73.31
320	7	90.39	67.11	31.80

Another, more realistic method to estimate the power of a linkage study is implemented in the ASP Power Calculator by Krawczak (2001) which uses a likelihood ratio test for allele sharing in sib pairs. Here, the incomplete information content of a marker and the recombination to the disease locus are

incorporated. However, the program is restricted to just four preset significance levels, and the power is calculated for an implicit two-point analysis. Power estimates are presented for the restricted likelihood ratio test, which is an extension of the mean test for not fully informative sib pairs and should therefore be comparable in terms of power with the NPL_{pairs} method from Genehunter (Kruglyak et al. 1996). Table 7 shows the results for different marker settings comparable to the marker closest to the disease locus in this simulation setup at α=0.01 under scenario 1. The difference to the theoretical estimates of maximal power is striking. Depending on the distance and the information content of the marker the power for 80 ASPs ranges from less than 20% to 42% instead of the maximal power of 55% at this significance level. Due to the loss of information, a bi-allelic marker (with equifrequent alleles) yields only about half the power of a six-allelic marker at the same locus (compare columns 1 and 2). Almost the same reduction can be seen if the marker is 7 instead of 1 cM away from the disease locus (see columns 1 and 3).

Table 7: Estimated power (in %) of a single study under scenario 1 (10000 simulations of ASP Power Calculator, restricted model) at a significance level of 0.01, for different distances between the disease locus and the closest marker locus and different numbers of equifrequent marker alleles.

Number of ASPs	1 cM, 6 alleles	1 cM, 2 alleles	7 cM, 6 alleles
40	21.1	9.6	12.2
80	41.7	18.9	24.2
320	97.1	71.0	82.2

Summarizing, the power is overestimated by Risch's formula as it assumes full information at the disease locus while the ASP Power Calculator is too pessimistic for multipoint analysis. The size of these deviations can be seen in the analysis of the individual simulated studies (table 8). The power is calculated as the fraction of simulations in which the p-value for the combined sample is below a given level, and thus the null hypothesis is correctly rejected. Here, the lowest p-value in an interval of 30 cM around the disease locus is considered. This is a practical compromise between counting any significant result on the whole chromosome and counting only the disease locus itself as a correct positive result. The first approach would be too optimistic as further investigations (e.g. fine mapping) are usually concentrated in a limited region around the most promising

finding. The latter approach would be too restrictive as it is well known that there is considerable variation in the location of the strongest result (Cordell 2001). This choice is also reasonable as GSMA can only localize a trait to within a bin (usually about 30 cM) and it would be unfair to require the other methods to localize it more precisely. If power is based on an interval of 14 cM around the disease locus, all results are between 0.2 to 2.6 percentage points lower (at the suggestive level), but with no clear pattern of differences between the meta-analysis methods.

Table 8: Power of single studies at different significance levels (percentage of replicates with a p-value below the respective level at a locus within 15 cM in both directions of the true disease locus, out of 1000 replicates) and average information content (IC) of each study at the disease locus under scenario 1.

Study type	$\alpha = 0.01$	$\alpha = 0.00074$	$\alpha = 0.000022$	Average IC
A	34.3±1.5	6.3±0.8	0.5±0.2	0.583
B	45.1±1.6	12.6±1.0	1.2±0.3	0.703
C	43.3±1.6	11.2±1.0	1.6±0.4	0.775
D	43.9±1.6	10.9±1.0	1.3±0.4	0.722
C_1	21.4±1.3	3.9±0.6	0.3±0.2	0.776
C_2	20.9±1.3	3.4±0.6	0.3±0.2	0.775
D_1	21.6±1.3	3.2±0.6	0.0±0.0	0.722
D_2	21.2±1.3	3.1±0.5	0.1±0.1	0.721

When there is less than perfect information (which is always the case), the NPL_{pairs} procedure is conservative under both the null and alternative hypotheses, but for different reasons. Under the null hypothesis, the variance of the test statistic is overestimated (Kong and Cox 1997), thus leading to a conservative test. Under the alternative (at the position of a disease gene) the test statistic will additionally be downward-biased, resulting in power loss. Thus low information in effect distorts the overall result in the direction of no linkage. This occurs because allele sharing is estimated given all marker data by using the distribution of the number of alleles identical by descent under the null hypothesis (Kruglyak et al. 1996). The expectation under the alternative of linkage would in general be higher. So in the extreme case when there is no information about the actual inheritance at a locus in a family, this is considered equivalent to the case when it can be observed that an ASP shares one allele identical by descent, which is what is expected under the null. But in the latter case there is actually have some

evidence against linkage, whereas in the former case there is in fact no evidence at all since nothing has been observed at all. This effect can be quite drastic. Figure 2 shows the information content (IC, as calculated by Genehunter (Kruglyak et al. 1996)) along the linked chromosome in one replicate of studies A and C each, both having markers with 6 equifrequent alleles, spaced 15 cM apart. The disease locus is highlighted. At the locations of the markers, the information content is almost 90%, and it drops to less than 60% between markers.

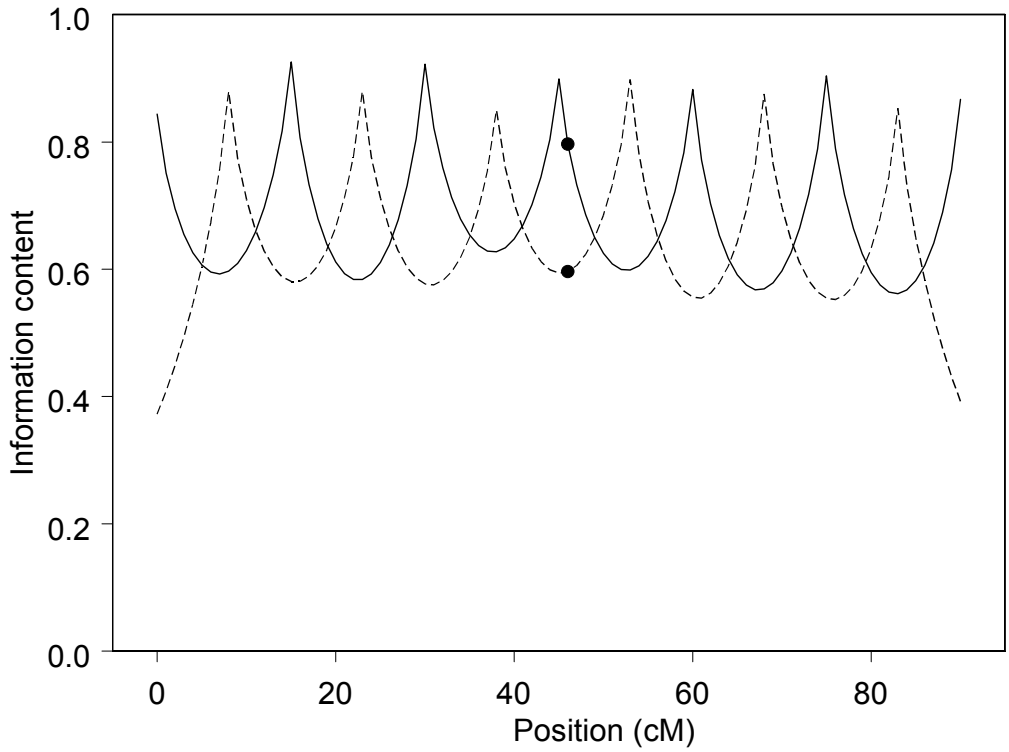


Figure 2: Information content along the linked chromosome for a replicate of study type A (dashed line) and C (solid line). The position of the disease locus is denoted by a dot.

The corresponding NPL_{pairs} statistics for this replicate are shown in figure 3. For study C, which has a marker close to the disease locus and therefore high information content at this position, the NPL_{pairs} score is 4.03 at the disease locus (p-value=0.000025) and 4.21 at the next marker (p-value=0.000011). For study A, on the other hand, the maximal score is only 3.0 (p-value=0.00297) at the flanking marker and 2.84 at the disease locus (p-value=0.00216).

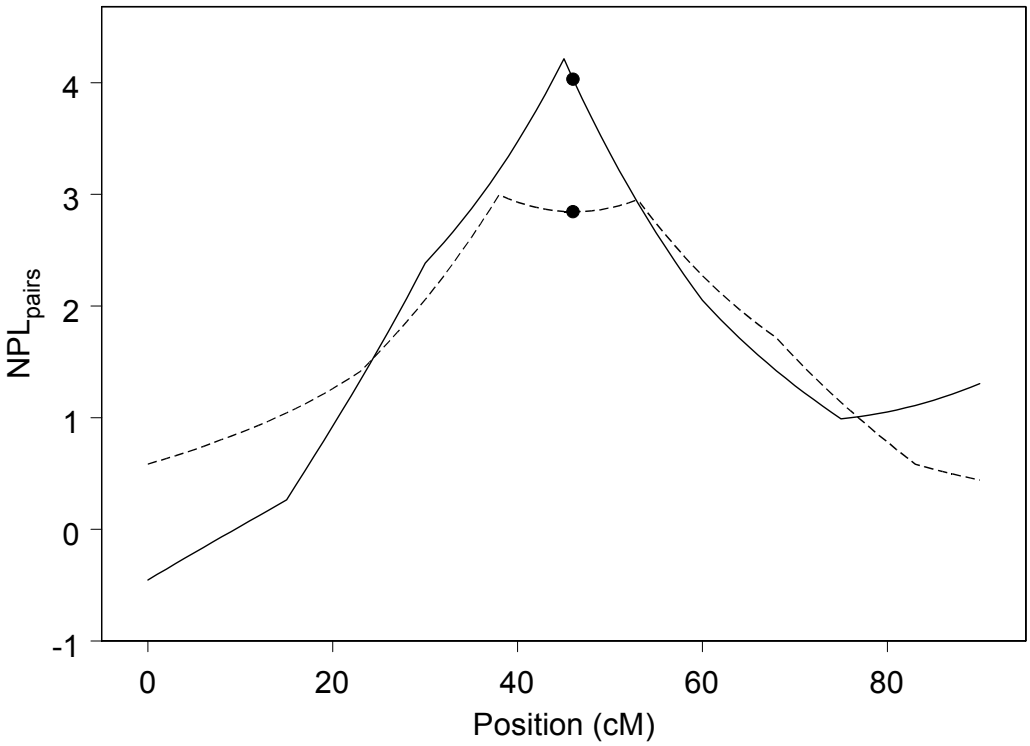


Figure 3: NPL Score along the linked chromosome for a replicate of study type A (dashed line) and C (solid line). The position of the disease locus is denoted by a dot.

Table 8 shows the power of the single studies of type A-D and their respective average information content at the disease locus in scenario 1. The distance between the disease locus and the closest marker can be seen in table 3. As expected, they all have considerably less power than under the assumption of full information but more power than estimated by Krawczak's method. At a suggestive α level, power estimates for 80 simulated sib pairs are between 6% and 13%, compared to 23% maximum power. As power for this level cannot be calculated by the ASP Power Calculator, the power at a significance level of 0.01 is also compared. Here, the maximum power would be 55%, estimates in this simulation study vary between 34% and 45%, while Krawczak's program gives 20% to 42%, and Chen and Deng's program calculates 30% to 51%.

For the microsatellite markers, the power is lower in study type A (where the closest marker is 7 cM from the disease locus and the average IC across families at the disease locus is 0.58) than in study type C (closest marker 1 cM from disease locus, average IC at disease locus is 0.78). The difference in power of

these two designs is due to the combined effect of less than perfect information and recombination between disease locus and the closest marker, which cannot be disentangled in this setup. The multipoint analyses of the SNP maps B and D have similar power and average IC at the disease locus, comparable to the microsatellite design C. Splitting the samples C and D into two studies of 40 pairs each decreases the power by more than half, at stringent α levels even to a quarter.

The power of single studies under scenarios 2 and 3 is given in table 9. As expected, scenario 2 has similar power to scenario 1 and scenario 3 has lower power.

Table 9: Power of single studies at different significance levels (percentage of replicates with a p-value below the respective level at a locus within 15 cM in both directions of the true disease locus, out of 1000 replicates) under scenarios 1 to 3 at a significance level of 0.00074.

Study type	Scenario 1: multiplicative, GRR = 4	Scenario 2: recessive, GRR = 4	Scenario 3: multiplicative, GRR = 3.5
A	6.3±0.8	6.3±0.8	4.2±0.6
B	12.6±1.0	10.0±0.9	4.1±0.6
C	11.2±1.0	11.8±1.0	5.4±0.7
D	10.9±1.0	10.5±1.0	4.8±0.7
C_1	3.9±0.6	3.8±0.6	2.3±0.5
C_2	3.4±0.6	3.6±0.6	1.7±0.4
D_1	3.2±0.6	3.6±0.6	1.3±0.4
D_2	3.1±0.5	3.9±0.6	1.5±0.4

4.2.2 Power of meta-analysis methods

The power of the different meta-analysis methods and weighting schemes to combine four studies of equal size (80 ASPs each) is shown in table 10. Again, it is always below the maximum power for 320 ASPs. The highest power, close to 80% at the suggestive level, is achieved by the Z-score combination methods proposed here. Among them, weighting with information content or distance parameter seems to have slightly better power at all levels. The p-value based methods (Fisher, TPM, GSMA) have around 65% power at the suggestive level. Here, the lowest power is estimated for Fisher's method, but at the level of significant linkage Fisher's procedure has higher power than TPM and GSMA.

Table 10: Power of combined analyses of four studies A-D (N=80 each, scenario 1, multiplicative, GRR=4) at different significance levels (percentage of replicates with a p-value below the respective level at a locus within 15 cM of the true disease locus out of 1000 replicates).

Method	$\alpha = 0.01$	$\alpha = 0.00074$	$\alpha = 0.000022$
$Z_{unweighted}$	96.6±0.6	77.3±1.3	37.5±1.5
Z_{IC}	97.0±0.5	79.8±1.3	39.7±1.5
Z_{Dist}	96.9±0.5	79.5±1.3	39.6±1.5
GSMA	87.8±1.0	62.8±1.5	26.2±1.4
Fisher	87.5±1.0	62.4±1.5	28.3±1.4
TPM	94.0±0.8	68.2±1.5	27.9±1.4

For the recessive scenario with a very similar effect size, the Z-score combination methods have a power of 70% (unweighted) to 73% (weighted with IC or distance) at a suggestive level, while Fisher's method and GSMA have about 57% and TPM 63% power (table 11). In a scenario with a smaller effect size (multiplicative with GRR of 3.5), the Z-score methods reach 42% to 46% power, while Fisher's method has 28%, GSMA 31% and TPM 33% power.

Table 11: Power of combined analyses of four studies A-D (N=80 each) at different significance levels for scenarios 2 and 3 (percentage of replicates with a p-value below the respective level at a locus within 15 cM of the true disease locus out of 1000 replicates).

Method	Scenario 2, recessive		Scenario 3, multiplicative, GRR 3.5	
	$\alpha = 0.00074$	$\alpha = 0.000022$	$\alpha = 0.00074$	$\alpha = 0.000022$
$Z_{unweighted}$	69.9±1.5	30.6±1.5	42.4±1.6	10.7±1.0
Z_{IC}	72.4±1.4	33.4±1.5	45.8±1.6	11.9±1.0
Z_{Dist}	73.3±1.4	33.8±1.5	46.1±1.6	12.2±1.0
GSMA	56.4±1.6	23.1±1.3	30.9±1.5	7.4±0.8
Fisher	57.0±1.6	22.7±1.3	28.5±1.4	6.9±0.8
TPM	62.6±1.5	23.7±1.3	33.4±1.5	6.5±0.8

The results for different study sizes are given in table 12. Here, the general picture is similar. One should keep in mind that, in each simulation, these are the same families as before, just split up differently, thus the power reached in table 12 has to be compared with that in table 10. However, only for the Z-score combination methods weighted additionally with the number of affecteds, the results are exactly the same (the small difference when weighting with IC occurs because IC is calculated on either 40 or the whole 80 families and thus differs

marginally). For all methods that do not account for study size, the power is slightly lower in the setting with the smaller samples. The Z-score methods disregarding sample size lose around 3 percentage points in power (in the following presentation differences in power will still be reported as percentage points using the % symbol for convenience). Fisher's method seems to be the least robust in this respect, the power drops from 62% to 53%. For TPM the difference is about 4%, for GSMA only about 2% (at the suggestive level).

Table 12: Power of combined analyses of six studies A-D2 (N=80 or 40) of scenario 1 (multiplicative, GRR=4) of different study size at different significance levels (percentage of replicates with a p-value below the respective level at a locus within 15 cM of the true disease locus out of 1000 replicates).

Method	**$\alpha = 0.01$**	**$\alpha = 0.00074$**	**$\alpha = 0.000022$**
$\mathbf{Z_{unweighted}}$	96.1±0.6	75.7±1.4	35.8±1.5
$\mathbf{Z_{Size}}$	96.6±0.6	77.3±1.3	37.5±1.5
$\mathbf{Z_{IC}}$	96.7±0.6	76.9±1.3	36.9±1.5
$\mathbf{Z_{IC+Size}}$	96.8±0.6	80.1±1.3	39.9±1.5
$\mathbf{Z_{Dist}}$	96.7±0.6	76.6±1.3	37.0±1.5
$Z_{Dist+Size}$	96.9±0.5	79.5±1.3	39.6±1.5
GSMA	87.5±1.0	60.5±1.5	24.1±1.4
Fisher	81.9±1.2	53.0±1.6	20.5±1.3
TPM	93.0±0.8	64.3±1.5	24.5±1.4

This picture is similar for the other simulated disease models (table 13), where the power of Fisher's method drops from 28% to 23% (multiplicative model with smaller effect) or from 57% to 49% (recessive model) at the suggestive level. The power of TPM drops by 3% and that of GSMA by about 1%, whereas the Z-score methods still have the same power as with four studies of equal size.

Table 13: Power of combined analyses of six studies A-D2 (N=80 or 40) at different significance levels for scenarios 2 and 3 (percentage of replicates with a p-value below the respective level at a locus within 15 cM of the true disease locus out of 1000 replicates).

Method	Scenario 2, recessive		Scenario 3, multiplicative, GRR 3.5	
	$\alpha = 0.00074$	$\alpha = 0.000022$	$\alpha = 0.00074$	$\alpha = 0.000022$
$Z_{unweighted}$	69.2±1.5	29.0±1.4	42.5±1.6	9.7±0.9
Z_{Size}	69.9±1.5	30.6±1.5	42.4±1.6	10.7±1.0
Z_{IC}	70.6±1.4	31.5±1.5	44.1±1.6	11.0±1.0
$Z_{IC+Size}$	72.3±1.4	33.6±1.5	46.0±1.6	11.8±1.0
Z_{Dist}	71.3±1.4	31.8±1.5	44.6±1.6	11.3±1.0
$Z_{Dist+Size}$	73.3±1.4	33.8±1.5	46.1±1.6	12.2±1.0
GSMA	55.3±1.6	19.1±1.2	30.7±1.5	7.5±0.8
Fisher	47.8±1.6	17.9±1.2	22.9±1.3	4.2±0.6
TPM	59.6±1.6	21.9±1.3	30.7±1.5	5.8±0.7

In a single replicate, there can be large differences for the p-value methods if the samples C and D are split or not, for instance in one replicate Fisher's method gave a p-value of 1 at the disease locus when six studies were combined and a p-value of 0.065 when the same families where arranged in four samples. For another replicate, the TPM method gave p-values of 0.0086 and 0.0659 at the disease locus, depending on the arrangement of families into subsamples. So, with the same data, the result can vary between significant to very far from significant. When a meta-analysis is classified as having a significant or not significant result at a certain level (again counting as significant a p-value below the significance level at any locus within 15 cM of the disease locus), the concordance between these results when arranging the families into six or four studies can be compared in a 2-by-2 table (with rows significant/not significant and columns 4/6 studies and one entry for each replicate). The Kappa coefficient of this concordance over the 1000 replicates is 1 for the weighted Z-scores, 0.84 for the TPM method, 0.71 for Fisher's method and 0.55 for GSMA on the 0.000022 level (on the 0.00074 level the Kappa for GSMA is 0.64, the other methods have the same Kappas as before). As can be seen, the advantage of the Z-score combination methods is even greater when studies differ in size.

4.2.3 Type 1 error (significance level)

The following presentation of the type 1 errors (tables 14 and 15) is intended to give only an indication of what can be expected with these meta-analysis methods. 1000 simulations are generally not enough to estimate the type 1 error rate with appropriate precision. Thus, only values to a pointwise level of 0.01 are reported in detail (for the other levels less than 3 false positives occurred with each method in 1000 simulations). These are pointwise significance levels which imply a different type 1 error in a genome wide analysis with multiple correlated markers. According to the formula by Lander & Schork (1994) in the simulated setup (one 90 cM chromosome) a pointwise α=0.01 corresponds to a chromosome wide α*=0.107, if an analysis with an infinitely dense marker map is assumed. Since this assumption is never fulfilled, the resulting correction from chromosome-wide to pointwise is too conservative. Thus the equivalent chromosome-wide level will in fact be considerably smaller.

The simulation of the single studies yields the chromosome-wide levels which correspond to the chosen pointwise levels in the chosen setup. For the single studies, chromosome wide false positive rates range between 0.022 and 0.045 (table 14), for the meta-analyses with p-value methods 0.013 to 0.033 and for the Z-score combination methods 0.016 to 0.051 (table 15). Thus, within the limit of precision attainable by 1000 simulations, the type 1 error rates seem to be acceptable. The three p-value methods might be slightly more conservative than the Z-score combination methods.

Table 14: False positive rates at a significance level of 0.01 (percentage of replicates with at least one p-value below 0.01 on an unlinked chromosome of length 90cM) for single studies.

Study type	Scenario 1: multiplicative, GRR = 4	Scenario 2: recessive, GRR = 4	Scenario 3: multiplicative, GRR = 3.5
A	3.4	2.5	4.2
B	3.4	4.3	4.3
C	3.6	4.3	4.1
D	4.0	3.0	3.5
C_1	2.2	3.9	2.7
C_2	4.5	3.8	4.4
D_1	4.0	2.7	4.5
D_2	3.6	2.9	3.7

Table 15: False positive rates at a significance level of 0.01 (percentage of replicates with a p-value below 0.01 for an unlinked chromosome of length 90cM) for combined analyses of four equally sized studies A-D (N=80 each) and of six studies A-D2 (N=80 or 40) of varying size.

Method	Four equally sized studies			Six studies, varying size		
	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
$Z_{unweighted}$	3.8	4.0	1.6	3.8	4.2	1.9
Z_{Size}	-	-	-	3.8	4.0	1.6
Z_{IC}	4.4	4.8	2.7	4.6	4.7	2.9
$Z_{IC+Size}$	-	-	-	4.6	5.0	2.6
Z_{Dist}	5.1	5.0	3.2	4.9	5.0	3.2
$Z_{Dist+Size}$	-	-	-	5.1	5.0	3.2
GSMA	2.2	2.6	1.3	2.6	3.3	1.9
Fisher	2.1	2.9	3.2	2.6	1.7	2.3
TPM	2.8	2.8	2.2	2.3	3.0	1.9

4.3 Discussion of the simulation study

This simulation study presents the first comparison of methodology for meta-analysis of genome wide linkage studies. In the simulated scenarios, under three different trait models with different genetic effect sizes, the p-value methods that have been applied so far in meta-analysis have consistently lower power than a direct combination of Z-scores. The unweighted combination of Z-scores gives 9% to 15% (9% to 13%) more power than the p-value methods at the suggestive (significant) level when combining studies of the same sample size (table 10 for the multiplicative model with GRR of 4). In the normal situation that studies are of different size, a weighting with size raises the power of the Z-score method by 2% to 3% compared to the unweighted method (table 12). And while this weighted method has the same power if the complete sample of families is split up into four or six studies, the p-value methods lose up to 9% power. And even more importantly, for a single replicate, the weighted Z-score combination gives the same result, while the p-value methods can give drastically different results on different partitions of the same data. In terms of power, Fisher seems to be the least robust, in terms of concordance of results, GSMA was the least robust when going from 4 to 6 studies with the same total sample.

When weighting with the additional information on marker map and information content as proposed here, the power increases by another 2% to 3%. Even though this gain is comparatively small, it is consistent across these simulations and as it comes at no additional cost, it should be exploited when the necessary information is available. The total difference in power between the best weighted Z-score combination and the p-value methods is 16% to 27% and 15% to 19% at the suggestive and significant level, respectively for the multiplicative model with GRR 4. For the recessive model it is 14% to 26% and for the multiplicative model with GRR 3.5 it is 13% to 15% at the suggestive level. The magnitude of possible gain in power depends on the specific situation, e.g. how much the studies differ in marker spacing and informativity. But in general, low information reduces power which can be interpreted as reducing the effective sample size of a study.

The false positive rates at a nominal 0.01 level were slightly higher for the Z-score methods than for the p-value methods but all very close to the expected level in this simulated setup.

The 15 cM microsatellite maps that were simulated represent gaps that are larger than the average marker spacing in genome scans, but it is common to have some gaps of at least this size (many published scans have about 10 cM average spacing, recent ones 3 to 5 cM but early ones even have 20 cM and all of these have to be included in a meta-analysis). And if the disease locus is in one of these gaps, the power to detect linkage is reduced substantially as could be seen in the individual studies. Situations with even worse constellations around the disease locus can occur by chance, e.g. when including candidate studies and sparser genome scans.

In the published applications of the p-value methods, study size and map information are usually ignored, even though appropriate weights could be integrated. However, in contrast to weighted Z-scores, the distribution of a weighted GSMA method under the null has to be derived by extensive simulations, which might be the reason why it is not performed in practice. For TPM and Fisher's p-value combination, a weighting of studies, especially with sample size, but extensible to other weights as well, has been proposed (Good 1955; Zaykin et al. 2002). It also relies on simulations under the null, since the exact distribution can be given only in special cases. Again, this improvement has not been widely adopted in applications of genetic epidemiology. For instance, Wu et al. (2002), and Badner & Gershon (2002a; 2002b) combine samples using modeifications of Fisher's method without accounting for their very different sizes. Other extensions of weighting schemes, such as for differences in family structures, ascertainment, or general aspects of study quality can be developed. The weights proposed here are a first approach to address some of the most common differences between studies. The weighting by distance to the closest marker has the advantage of being independent of the observed genotype data and could be further refined by including a measure of the marker informativity such as polymorphism information content (PIC, Botstein et al. 1980), linkage information content (LIC, Guo and Elston 1999; Guo et al. 2002) or multipoint PIC value (MPIC, Rijsdijk and Sham 2002).

As noted above, it seems that the deviation under the alternative model is better captured by the Z-scores than by p-values, independent of the weighting. As a standard multipoint analysis often includes the calculation of NPL-scores, combining the Z-scores instead of p-values would be straightforward. The

advantage of the Z-score method can also be utilized for genome scan data analyzed with any other linkage statistic, as Z-scores can be derived from p-values by the inverse of the normal distribution. This procedure was apparently introduced by Stouffer (1949, as cited by Hedges and Olkin 1985) and is also known as the inverse normal method. In general, caution must be applied when combining results from different test statistics - p-values as well as Z-scores - that have different underlying assumptions on the disease model.

In their article introducing TPM, Zaykin et al. (2002) compare their method with Fisher's and the inverse normal method, among others. They directly simulate Z-scores under the null and under a shift alternative and investigate the situation where only some of the null hypotheses are false. This corresponds to genetic heterogeneity between studies, where only in some samples a disease gene segregates at the studied position. In this case, the TPM method performed best, and the power of Fisher's procedure is higher than that of the inverse normal method. When all null hypotheses are false, i.e. the same disease locus is present in all studies, the inverse normal method has much higher power than TPM. For this simulation study, data was simulated under the same alternative for all studies, i.e. under the same disease model for all studies and thus similar results were obtained. The power of a meta-analysis compared to individual studies is highest for genes common to all investigated populations - for population specific genes the power necessarily drops. TPM accounts for possible genetic heterogeneity and therefore is less efficient when this is not present. Optimal methods for combining statistical tests depend on the alternative, especially whether all or just few of the null hypotheses are false. Some theoretical work can be found in a review article by Folks (1984), who investigates in which situations the inverse normal method, Fisher's method or truncation methods are better. Further simulations for other realistic models of gene effects and heterogeneity within or between studies might be of interest, since a generalization of results always depends on the simulated model. The methodology investigated here was developed for studies that supposedly are homogeneous in this respect and simulations were done accordingly. The proposed weighting schemes correspond to fixed effects meta-analysis (Hedges and Olkin 1985). To account for genetic heterogeneity between studies, a random effects meta-analysis could be more appropriate (e.g. Li and Rao 1996). If the average number of alleles shared

identically by descent between affected sib pairs is the parameter of interest, the estimated numbers of sib pairs sharing 0, 1 or 2 alleles are necessary to calculate the appropriate between study variance of the parameter estimates (Gu et al. 2001). However, most commonly used linkage software programs do not output the necessary statistics which greatly hampers such a random effects analysis. McQueen et al. (2006) recently proposed a bootstrap variance estimator for the average number of alleles shared identically by descent between affected sib pairs to be used in a random effects meta-analysis. Again, this relies on the availability of detailed data, here the family specific IBD estimates, which can be output by analysis programs such as Genehunter (Kruglyak et al. 1996) or Merlin (Abecasis et al. 2002), but are usually only available if the raw genotype data are also available. NPL scores and information content on the other hand are normal output of linkage programs for affected relative pairs and genome-wide NPL scores are usually also published. If conducting a meta-analysis with studies from very diverse populations it might be advisable to first combine all samples to find common genes and additionally analyze combinations of more closely related subsets to identify population specific genes (as done by Wu et al. 2002).

5 Application of the Meta-Analysis Method: the GAW12 Asthma Data

Identifying susceptibility loci for complex diseases such as asthma with the use of genome scans is a difficult task. One family study alone seldom yields statistically significant results for susceptibility genes of moderate effect. Higher power can be expected with a pooled linkage analysis of the combined raw data or a meta-analysis of outcomes from different genome scans. Methodological problems are how to account for differences in population, ascertainment and family structure, phenotype definition and marker sets between studies. Weighting schemes for combining results from several genome scans are proposed. The GAW 12 asthma studies vary, partly extremly, in sample size, marker density and information content at each chromosomal location. For the combined analysis a simple sum or average of individual scores cannot account for these differences. To avoid confounding, different approaches of weighting the scores accordingly are investigated.

For a genome scan using multipoint linkage analysis of the pooled data sets, a common marker map is necessary, consisting of all markers genotyped in any one of the studies. The GAW 12 studies use at least partly different markers. From existing, publicly available marker maps the order and distances of all markers could not be determined unambiguously. Multipoint linkage analysis with a dense marker set has been reported as sensitive to misspecification of marker order and intermarker distance (Halpern and Whittemore 1999). Two different maps were constructed and used for multipoint linkage analysis to compare the results for discrepancies attributable to map differences.

5.1 Data sets

The data of GAW12 problem 1 consists of four genome scan and five chromosome 5 data sets: the genome scans from the Collaborative Study on the

Genetics of Asthma (CSGA, Xu et al. 2001), the Hutterites (Ober et al. 2000), Germany (Wjst et al. 1999), and Busselton in Australia (Daniels et al. 1996). The chromosome 5 data sets are from the Consortium on Asthma Genetics (COAG, Lonjou et al. 2000) and come from Southampton, Finland, Perth, Oxford and Freiburg as described by Palmer et al. (Palmer et al. 2001). The binary phenotype asthma was considered and all pedigrees with relevant affecteds for linkage analysis were used (see table 16), i.e. those with at least two affected genotyped members excluding parent-child pairs. Because of pedigree size limitations for multipoint linkage algorithms one pedigree (No.97) of the Oxford sample had to be divide into two families and the Hutterite pedigree was split into 14 families with at least two affected genotyped members by discarding most of the unaffected and untyped individuals.

Table 16: Families with at least 2 asthmatics relevant for linkage analysis in the GAW12 data sets.

Study	CSGA	COAG Southampton	Germany	COAG Finland	COAG Perth	COAG Oxford	Hutterite	COAG Freiburg	Busselton
Families	225	88	97	52	43	30	14	14	4
Relevant Affecteds	596	214	200	122	115	81	48	30	8
Subjects	1492	517	415	399	235	222	174	70	20

5.2 Methods

5.2.1 Common Marker Map Construction

A common marker map that contains all markers typed in any study was first constructed. This map is based on the Marshfield comprehensive human genetic linkage map (Broman et al. 1998) and was completed for markers without Marshfield distance by using genotype information of the GAW 12 studies. To assess sensitivity of the analysis method to marker order and intermarker distances, two maps that differed slightly in marker order and considerably in intermarker distances were assembled.

For each study marker name and order was provided. Marker distance (cM from p-ter) was only provided for the Hutterite and CSGA data sets. There were a total of about 900 markers across studies.

The problems for the construction of the common map were:

- Choice of a reference map.
- Location assignment of markers used in a study but missing in the reference map.
- Handling of different markers located at the same position.

Two different combined maps resulted, which are based on the same reference map, but with different handling of location assignment and same positions.

1. Reference map: The sex-averaged Marshfield comprehensive human genetic linkage map (http:// research.marshfieldclinic.org/genetics/, Broman et al. 1998) was used as the reference map. For an entered list of markers corresponding distances are given, which is especially helpful for a large number of requested marker locations. The locations of typed markers in the Hutterite study correspond to the Marshfield map (rounded to one cM). If available, distances given by the CSGA or Hutterite study descriptions were used for markers missing in the Marshfield map.

2. Location assignment: Markers not appearing in the reference map were assigned a location in two different ways yielding slightly different maps. Given the GAW 12 data sets the *order-based map* utilizes only the order of markers and the *data-based map* additionally incorporates genotype information. For the order-based map, missing markers were assigned a position corresponding to the provided order and equidistant to the flanking markers of the Marshfield map. For the data-based map, the location were assigned by two point and multipoint linkage analysis using the MLINK and LINKMAP options of LINKAGE (Lathrop et al. 1984) and the available genotype data. Based on the given marker order, linkage analyses with the closest typed markers and their Marshfield positions were performed. When several adjacent markers were missing, their relative positions were assigned by combining the results for these markers relative to each other and with neighboring markers of the Marshfield map.

3. Same position: For multipoint linkage analysis different markers have to be assigned distinct positions. In the Marshfield map markers separated by little or no genetic distance quite often have no recombination events in the CEPH families used and therefore are presented with the same position in arbitrary order. Markers from different studies with identical Marshfield map positions were

merged to one artificial marker in both maps. For the order-based map markers analyzed in the same study with the same position in the Marshfield map were set 0.01 cM apart with their order as given in the study. For the data-based map they were tested for recombinations using two point linkage analysis. If no recombination was observed, the same procedure as in the order-based map was used. Otherwise markers were located by multipoint linkage analysis.

5.2.2 Linkage analysis and weighting schemes for meta-analysis

Multipoint linkage analysis of each study was performed with the ALLEGRO program (Gudbjartsson et al. 2000) an improved version of Genehunter (Kruglyak et al. 1996). For each pedigree the nonparametric linkage score statistic Z_{lr} (using the exponential model as recommended by Kong and Cox 1997) was calculated at each position where a marker was typed in any of the studies. This score is based on IBD sharing in affected relative pairs. The pointwise scores were averaged using different study specific weights. Multipoint linkage analysis gives scores for every position along the genome and thus enables the combination of results from studies in which different marker sets were analyzed. However, when adding scores for each position across studies the differences in study size (number of affecteds) as well as the information content of each study at that position (or the distance to the next typed marker as an alternative measure of information content) should be taken into account so as not to bias the results towards the null. This can be done by using appropriate weights on the study scores.

The following weighted and standardized statistics were used (as defined in chapter 3):

Z_1: equal weights

Z_a: weights relative to the number of relevant affecteds included in the study

Z_ic: weights relative to the number of relevant affecteds and to the information content of the considered locus in the specific study. As a measure for information content, the per-family information measure calculated in ALLEGRO was used.

Z_d2, Z_d10: weights relative to the number of relevant affecteds and an exponential function of the distance of this locus to markers typed in that study. The distance to the closest typed marker was calculated and used as argument of the exponential function with parameter -1/2 or -1/10.

These weighting schemes were compared with the Genome Search Meta-Analysis Method (*GSMA*) proposed by Wise et al. (1999). As described in chapter 2, the chromosomes are split into bins of approximately equal length for GSMA. A bin width of approximately about 30 cM was used as proposed by the authors, resulting in 113 bins. For each study, the most significant result in each bin is recorded. The bins within one study are ranked accordingly. Then the ranks for each bin are summed across all studies. The exact distribution of the summed ranks under the null hypothesis of no susceptibility locus in a specific bin is given by the authors. The markers in the Busselton study are spread rather unevenly over the genome resulting in six bins containing no Busselton marker. One other bin contained no CSGA marker. Even increasing the bin size from the recommended 30 cM to 40 cM left some empty bins and created the additional problem on chromosomes 21 and 22 of using either one bin considerably larger or two bins considerably smaller than all other bins. Therefore 30 cM bins were used and only the ranks 1 to 53 and 60 to 113 were assigned to the Busselton data. Rank-sums and p-values for those bins were calculated as if there were only three studies. GSMA can only be used for the four genome scans and is not valid for the COAG data of the chromosome 5 candidate region. Application of GSMA to only a candidate region would violate the assumption of at least one marker in each bin and the distribution of the ranks of the candidate region alone would differ from their distribution among ranks in the whole genome. At GAW12 Wise presents an extension of GSMA to this situation where the null distribution is approximated by simulation (Wise 2001).

5.3 Results

5.3.1 Common Marker Map Construction

For a total of about 800 markers the map locations could be assigned by the Marshfield linkage map. For about 100 markers the position was missing or more than one marker was at the same position in the Marshfield map.

The data-based and order-based maps contain 832 or 828 markers, respectively. Some markers merged into one marker in the order-based map could be assigned different positions in the data-based map. One marker (D3S11), allegedly located on chromosome 3 in one study, was positioned on chromosome 2 (for both maps) by the linkage analysis using the genotype data of this study in

agreement with several other physical and linkage maps. The data-based and order-based maps differ as follows:

1. Differences in location assignment: 47 markers were analyzed in at least one study, but were not contained in the Marshfield map or in the CSGA and Hutterite study description. These markers are located at different positions in the two constructed maps, the differences amounting to 1 cM for 13 markers, 2-4 cM for 16 markers, 5-10 cM for 14 markers and more than 10 cM (11-23 cM) for 4 markers.

2. Differences for same positions: For a total set of 52 markers, two or more were located at the same position in the Marshfield map and genotyped in a single study. For 13 out of these 52 markers the positions in the two constructed maps were identical, since linkage analysis did not show any recombinations. For 39 markers linkage analysis resulted in positions which differed from the original reference map. The localization of these 39 markers in the two constructed maps differs by 1 cM in 15 cases, by 2-4 cM in 13 cases, by 5-10 cM in 9 cases and 15 cM in two cases.

To compare the results of the different maps, the analysis where Z-scores for each study were weighted relative to the number of relevant affecteds in the study and the information content calculated by ALLEGRO is considered. The results of this linkage analysis are presented as p-values yielded by the overall score statistic for each genome position. In general, differences in p-values for the two constructed maps were negligible, even where markers were differently positioned. There are a few regions for which small differences in p-values can be noted. Figure 4 shows the results for the region with the largest differences across the whole genome which was on chromosome 5 (130-156 cM). Figure 5 shows the corresponding maps.

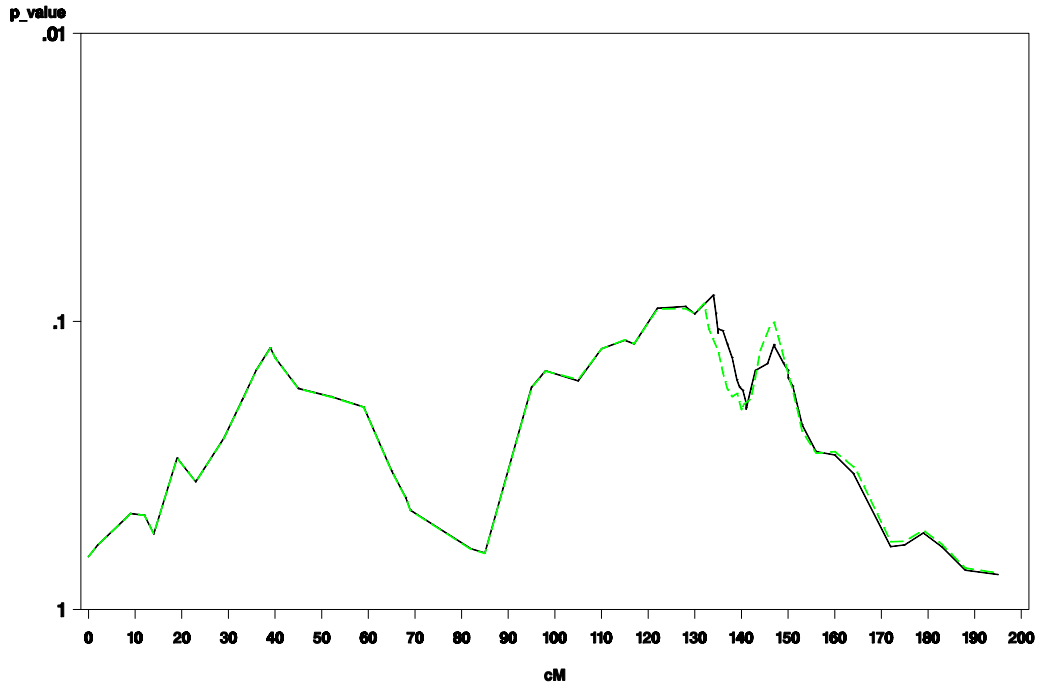


Figure 4: Multipoint linkage results using ALLEGRO for chromosome 5 based on the data-based map (solid line) and the order-based map (dotted line).

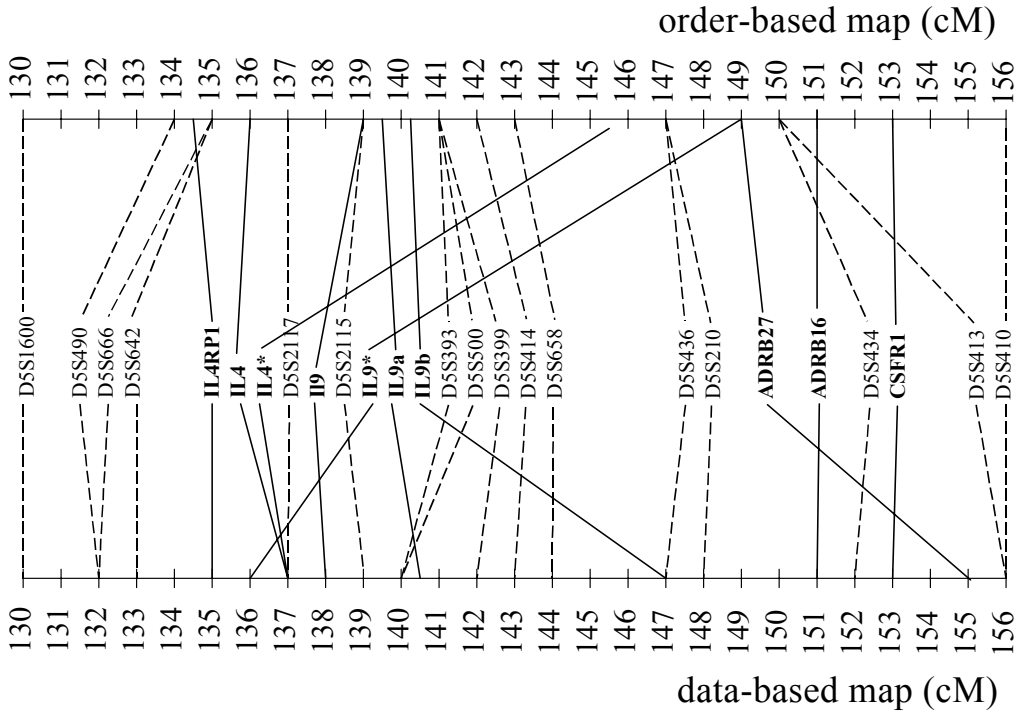


Figure 5: Chromosome 5, region 130-156 cM. Presented are markers which are in this region for both maps.; * indicates discrepancies for the same gene due to different data descriptions.

The maximal difference is at position 137 cM with p-values 0.172 and 0.120 for the data-based and order-based map, respectively. There were three positions (all within 3 cM) with differences in p-values greater than 0.04. Considering individual markers, the differences in p-values are maximal at marker D5S413 (data-based map: 156 cM, p-value 0.287; order-based map: 150.01 cM, p-value 0.157) and marker ADRB27 (data-based map: 155 cM, p-value 0.271; order-based map: 149 cM, p-value 0.140). There were eight more markers in this region with differences over 0.05.

5.3.2 Linkage analysis and weighting schemes for meta-analysis

In the following the results of the combined analyses (figure 6, rows 1 – 6) are elucidated and compared to those of the single studies (figure 6, rows 7 – 10) to assess the weighting schemes. Here the data-based map was used as a common marker map.

GSMA: GSMA gives by design only one p-value per bin, none below 0.01. The lowest p-value is attained for the last bin of chromosome 11.

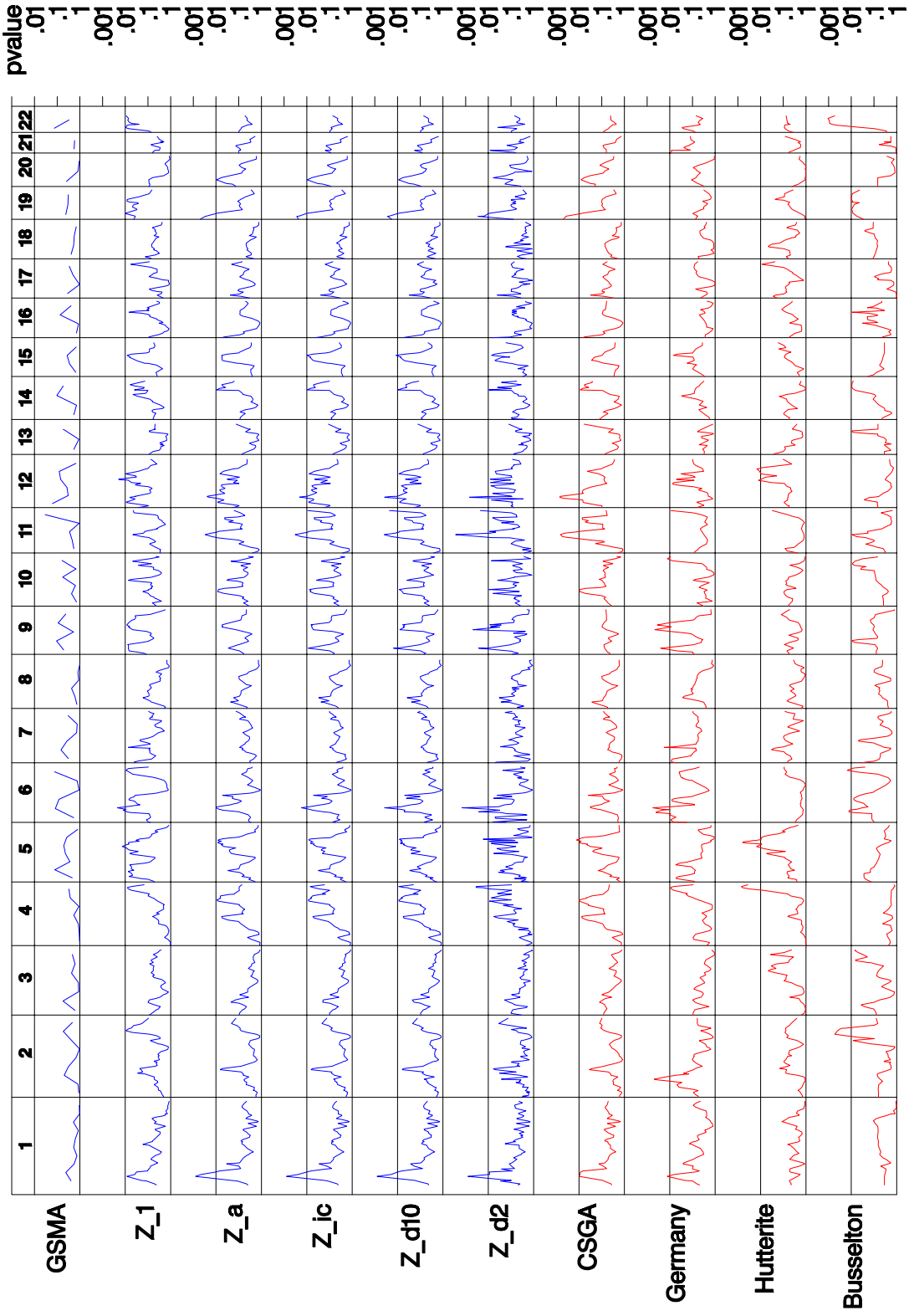


Figure 6: P-values along the whole genome for GSMA (row 1), the different combined scores Z_1, Z_a, Z_ic, Z_d2, Z_d10 (rows 2 - 6) and each genome wide study (rows 7 – 10); Note: inverse logarithmic scale with a line indicating a screening level of 0.01.

Z_1: Of the combined scores, the unweighted Z_1 is closest to GSMA. Again, the p-values are far from significant and the minimal values are higher than those obtained for the single studies. However, the combined minimal p-values are considerably smaller than those obtained by GSMA: below 0.01 at D5S421, 14cM away from the smallest value for the Hutterites, around D6S291, reflecting the German result supported by Busselton data, and at D12S327, combining the low values of German and Hutterite data, while the lowest values for the Hutterites 35cM away are leveled out by the other studies.

Z_a: When each study is weighted by the number of relevant affecteds slightly lower p-values were observed for the combined score at regions distinct from those indicated by Z_1. On chromosome 1 the low p-value of the German study of 0.009 corresponds to 0.002 in the combined analysis. The p-value 0.003 of the CSGA study at the marker BETA corresponds to a p-value of 0.006 in the combined analysis, for D19S886 the p-value of 0.004 remains, while the peak on chromosome 12 is slightly shifted. Scores from the larger studies (CSGA, Germany) have considerably more weight. Thus small p-values from these studies are hardly influenced by other studies.

Z_ic: Use of the information content and the number of relevant affecteds gives a similar picture. Because of large differences in study size, the information content has a relatively small impact on the weights but close examination shows small differences between Z_ic and Z_a in p-values and peak locations in most regions.

Z_d10: The weighted score Z_d10 again shows much similarity to Z_a because the map density has only limited influence on the weights compared to the number of relevant affecteds per study. The agreement with Z_ic is even greater, reflecting the correspondence of data driven information content and theoretically chosen distance measure. However, p-values differ to some degree as for example Z_d10 gives a p-value below 0.01 for chromosome 9 which Z_a and Z_ic do not.

Z_d2: The results for the weighted score Z_d2 show much more variation. For markers untyped in a study the distance measure can lower the weights so that the combined score could essentially be based on a single or two smaller studies only. P-values below 0.01 were obtained on chromosomes 4 and 5 and in the same

regions as indicated by Z_ic on chromosomes 1, 6, 9, 11, 19 . A suggestive result was reached at the marker BETA with a p-value of 0.0006.

For all methods the single p-values below 0.01 at D7S528 and D10S1248 for the German data, the distinct regions for German and Busselton data on chromosome 2, and the broad region indicated by Busselton data on chromosome 22 are leveled out by the other studies.

Figure 7 gives the combined results and individual weighting functions for all nine studies on chromosome 5. To investigate the direct impact of the weighting functions, the individual study scores Z_{lr} need to be considered. In addition to the number of relevant affecteds, the information content or the distance function, respectively, calculated for each study and each marker yields the basis for the different weighting functions. Figure 7 displays the results for chromosome 5 for individual studies ordered in descending order of the number of relevant affecteds (see table 16). The first row shows the p-values for the indicated scores in the combined analysis. Among the remaining rows the first column shows the p-values for the score Z_{lr} for each individual study. The next columns, denoted by Z_ic, Z_d10, Z_d2 present the information content and the distance function, respectively, on a scale from 0 to 1 which yield the weighting functions for the corresponding score.

There is basically zero weight for the COAG studies outside the candidate region and very high weight within. All studies show high weights within the candidate region. The weights for Z_ic are smoother than the ones incorporating marker distance. Z_d2 displays gaps between markers with basically zero weights especially for the genome scan data. Z_d10 displays a reduction of weights in marker gaps rather similar to Z_ic, with higher weights than Z_ic at the typed markers. Z_d10 seems to give higher influence to a wider candidate region than Z_ic. Note that the information content and distances measures given in figure 7 will be weighted again by the number of relevant affecteds such that their influence decreases rapidly with descending order in the figure.

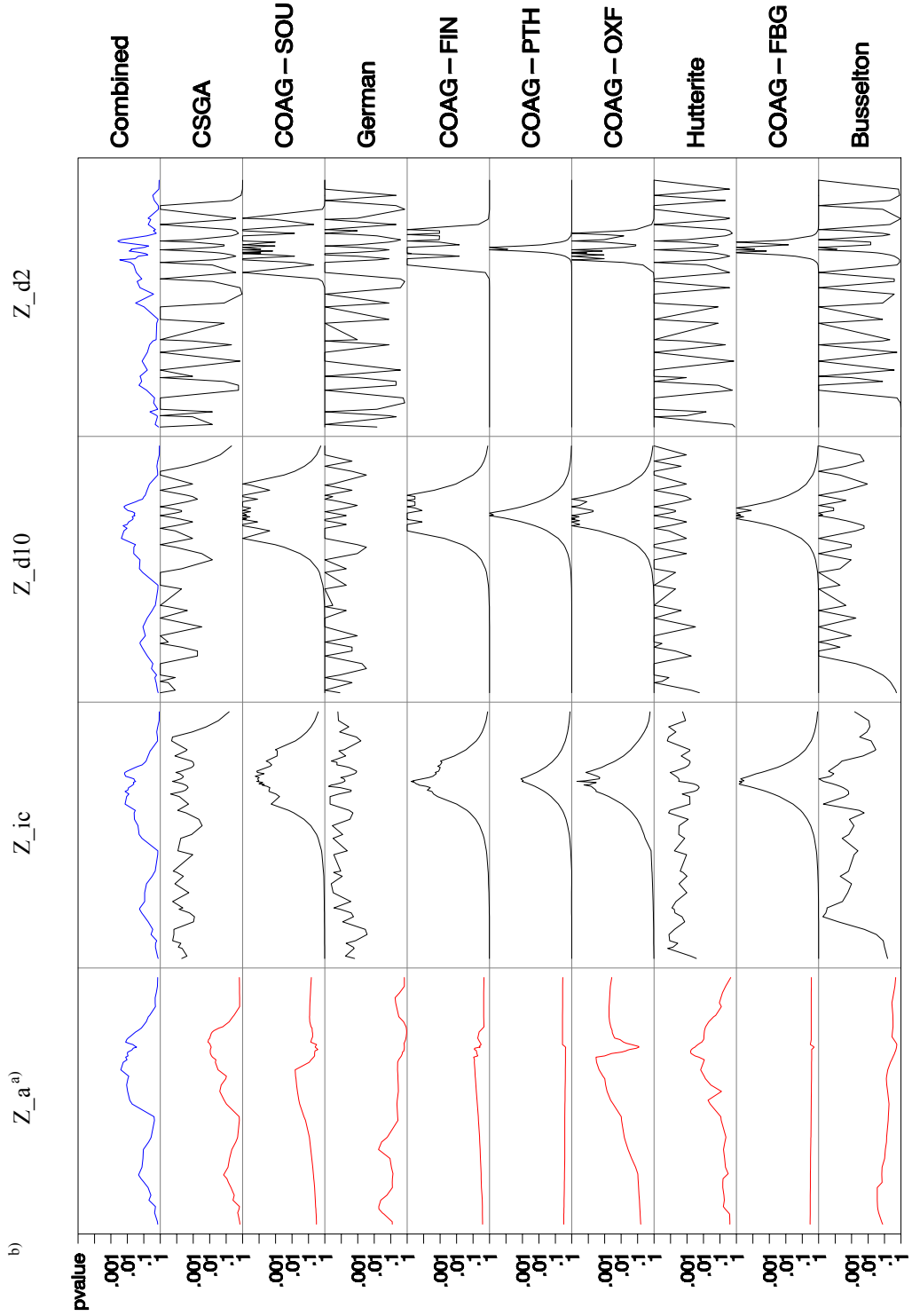


Figure 7: Combined and individual p-values, information content and distance measure for chromosome 5. Row 1: p-values for the indicated combined scores, remaining rows: [a)] Column 1: individual p-values of the single studies, column 2: information content, column 3&4 distance measure on which weighting functions for indicated scores are based ([b)] scale 0-1 for weights, omitted); Map distance omitted.

5.4 Discussion

Differences in the linkage results based on the two constructed maps must be due to differently placed markers. Hence, for the most part of the genome differences were expected to be negligible. However, even around the differently placed markers, only a few differences could be found with a large effect on the absolute p-value. No regions with p-values under a screening level of 0.01 could be identified in this combined analysis. Thus in the current situation, none of the differences due to the marker maps are relevant for the interpretation of results. The largest differences can be seen for a candidate region on chromosome 5 which was available for all data sets, including very small studies. Here, the differences in the marker maps are largest, since the Marshfield map, which contains only anonymous markers but not candidate genes, did not contain many markers of this region.

Overall are these asthma linkage results not sensitive in a relevant manner to the differences in the maps. Some large differences in p-values however indicate that map differences could change results for linkage regions between significant, suggestive and even non-significant. Previous studies have shown that serious map misspecification can result in appreciable effects on power and false positive rates of multipoint linkage analysis (Halpern and Whittemore 1999; Daw et al. 2000). Since the availability of the human genome sequence as a physical map (Lander et al. 2001; Venter et al. 2001), marker order should not be a relevant problem for linkage analysis any more. Additionally, more precise linkage maps have also been published (Kong et al. 2002) and recently, improved databases which combine information from linkage and physical maps have been published (Kong et al. 2004; Duffy 2006).

This analysis, which does not specifically incorporate heterogeneity of data sets, shows that suggestive regions previously identified in single data sets could no longer be identified in a combined analysis. This might be due to shifts in estimated linkage location for individual data sets. A heterogeneity test for these genome scans has been done by another group at GAW 12 (Babron et al. 2001), and it was concluded, that the studies do not show significant evidence for heterogeneity. However, the authors caution that this test probably has low power for just four different genome scan samples.

An important reason for the lack of significant results may be that the combined sample size is still too small to identify linkage to a complex trait such as asthma. The whole genome scans included only 340 families, for the chromosome 5 candidate region 567 families were available. Given the moderate power of ASP linkage analysis for complex diseases, this seems to be insufficient to reliably identify genes involved in asthma susceptibility. The only asthma susceptibility gene that was by now identified by positional cloning, the ADAM33 gene on chromosome 20p13 (Van Eerdewegh et al. 2002), was identified in a genome scan of 460 families with a LOD score of 2.94 and through the help of subsequent association studies. Linkage to this genetic region has not been observed in most other asthma genome scans, including those in this meta-analysis, which gives a minimum p-value of 0.01 on chromosome 20. In further association studies with larger sample sizes the effect of this gene appears smaller than in the initial report (especially in other populations, Blakey et al. 2005; Holgate et al. 2006), probably due to the expected bias in effect estimates from initial studies for loci identified through genome wide linkage analysis (Göring et al. 2001). A meta-analysis of association studies estimated the odd-ratio as 1.46 for a SNP in this gene. Therefore, this linkage meta-analysis of only 340 ASP families did not have sufficient power to identify this gene and probably also had low power for other asthma susceptibility genes.

The different weighting schemes are based on both the number of relevant affecteds and on the corresponding information content or distance measure in each study. The number of relevant affecteds is highly variable between the studies (range 8-596) and thus has overwhelming effect on the weighting functions. Family structures varied also considerably between studies (affected sib pairs to extended pedigrees) and it should be remarked that more distant affected relative pairs (such as e.g. second cousins) offer greater power for linkage analysis than close relative pairs such as siblings. Additionally the power varies whether connecting relatives in the pedigree are genotyped or not and a useful definition of the relevant sample size in a linkage study depends also on the specific linkage test statistic used. Taking the number of affecteds without regard for the specific family structures as the weighting factor is a compromise which approximates the relevant sample size for linkage analysis in this setting (for a more thorough discussion of the issue of relevant sample size, see section 6.3.4 below). The

results of three large studies (CSGA, Germany and for chromosome 5 COAG Southampton) dominate the combined scores, except if the variation in marker density is given extreme weight.

The information content measure depends not only on the informativeness of the observed genotypes but also on the family structure and density of the typed markers so that the weighting scheme for Z_ic implicitly incorporates all these factors. The weighting according to the distance of typed markers for Z_d2 and Z_d10 on the other hand depends only on the used markers but not on the actual data. Z_d2 might be too variable for practical use and only sensible in combination with a smoothing procedure.

The GSMA method seems to have comparatively low power. To some extend this can be explained by the disregarding of differences in the size of the studies and the set of markers used. Still, compared to the unweighted combined scores GSMA p-values are higher. A problem with the GSMA method arises when typed markers are not evenly spaced and covering each chromosome with appropriate bins can become impossible. The extension of GSMA to include studies covering only candidate regions or to incorporate weights involves extensive analysis of simulated data which might be too time-consuming for multipoint analysis of larger families. GSMA is a true meta-analytic approach adequate for published data, but a combined analysis using as much information of the data as possible is preferable.

Pooling or meta-analysis is the only possibility to substantially increase the power of genome scans to identify genes of moderate effect. Results of the pooled analysis of the GAW 12 asthma data as a real data application of the proposed methods are promising although only one problem of pooled analysis, the use of different marker sets, was addressed. The appropriate weighting of family data with large differences in marker sets is necessary to avoid bias of the results towards the null. The proposed methods with weighting schemes, which do not require the raw data but only scores, measures for weights and a common map, seem feasible for multi-center analysis and collaborative studies.

6 Application to the Quantitative Phenotype Adult Height

Many epidemiological studies or genome scans for common diseases come up with large and well characterized samples. If a sufficient number of the recruited individuals are related and additionally DNA or genotype information is available, linkage analysis for several traits can be conducted. This was done in 12 publications reporting genome scans in 28 separate samples for linkage with adult height (Deng et al. 2002; Geller et al. 2003; Hirschhorn et al. 2001; Perola et al. 2001; Thompson et al. 1995; Wiltshire et al. 2002; Wu et al. 2003; Xu et al. 2002; Sale et al. 2005; Sammalisto et al. 2005; Willemsen et al. 2004; Liu et al. 2004). Most of these were performed in samples ascertained for specific diseases unrelated to body height such as diabetes (Wiltshire et al. 2002) or asthma (Wu et al. 2003) while a few were performed in population samples such as the Framingham Heart Study (Geller et al. 2003).

Adult height (stature) is a highly heritable trait, with heritability estimates around 0.8 (Preece 1996; Silventoinen et al. 2000; Silventoinen 2003; Xu et al. 2002). In most of the published genome scans, heritability was also estimated from the study sample and reported values are between 0.69 (Perola et al. 2001) and 0.98 (Wu et al. 2003) (see table 19 below for the individual estimates). Height is a trait that follows a normal distribution in the whole population, as noted by Pearson and Lee already 100 years ago (Pearson and Lee 1903), who explained how this can be the result of many genes, each with a small, additive effect independent of the others, termed a polygenic model. But more recent segregation analyses showed also evidence for major genes on top of the purely polygenic inheritance (Ginsburg et al. 1998; Xu et al. 2002).

Many candidate genes for stature and growth-related traits have been proposed and studied for association. These include genes of the growth hormone-IGF-system (e.g. *GH1, GHR, GHRHR, GHSR, IGF-1, IR, STAT5b*), genes regulating bone formation (e.g. *COL1A1, BMP2, FGFR3, VDR*), genes involved

in pituitary development (e.g. *POU1F1, PROP1, LHX3, LHX4, HESX1*), and several others (Kant et al. 2003, see also table 25 in the discussion). Studies relating to these candidate genes and monogenic forms of growth disorders have been extensively reviewed (e.g. by Palmert and Hirschhorn 2003).

Additionally, the genetic analysis of variation in stature can be supplemented by the study of genetic syndromes which include short or tall stature among their cardinal features, such as Noonan Syndrome (OMIM 163950), Prader-Willi Syndrome (OMIM 176270) and many others. In some cases, genes responsible for these syndromes might also have alleles which influence normal growth variation or the whole syndrome is caused by microdeletions which include dozens of genes (e.g. Prader-Willi Syndrome), just one of which might be involved in growth regulation. Similarly, the short stature seen in Léri-Weill dyschondrosteosis (LWD, OMIM 127300) and Ullrich-Turner Syndrome is caused by haploinsufficiency due to heterozygous deletions or mutations of the SHOX gene or its regulatory regions (Attie 2000; Benito-Sanz et al. 2005; Rao et al. 1997). Such mutations or deletions in the SHOX gene seem to be responsible for a fraction of patients with idiopathic short stature as well.

This meta-analysis of linkage genome scans for adult height was conceived when analysing a genome scan of the Framingham Heart Study for this phenotype and noting remarkable overlapping linakge peaks with some of the previously published genome scans for height. However, there were also other scans which had identified different regions that did not seem to be replicated by our own results. Therfore a meta-analysis of all available genome scans for the phenotype stature seemed highly desirable to exactly quantify the evidence for linkage in these regions.

In this chapter, the linkage scan for height in the Framingham Heart Study sample is presented first, followed by the meta-analysis of genome scans for adult height.

6.1 Framingham Heart Study genome scan for height

The Framingham Heart Study is a large epidemiological cohort study, started to investigate risk factors for coronary disease. The sampling approach for the original cohort, which was recruited on a household basis, i.e. all household members above a certain age where asked to participate, lead to many related

participants which where extended for the Framingham family study with an offspring cohort in 1971.

GAW13 provided genetic and anthropometrical data from 330 general pedigrees of the Framingham Heart Study. In this sample, the genetics of height were studied using a two-stage approach, which ensures that all individuals can be analyzed together. First, regression models for the phenotypes were built to obtain a single adjusted trait value for each individual. At the second stage, a linkage analysis incorporating all genotyped individuals was performed.

6.1.1 Methods

6.1.1.1 Study group

The individuals from the Framingham Heart Study were recruited at two time points (the original cohort in 1948 and the offspring cohort in 1971) from the general population excluding those with cardiovascular diseases, heart attack, or stroke. Almost all participants were of Caucasian origin. From the 330 largest pedigrees with 4692 members, DNA was available for 1702 individuals, who were genotyped for 401 markers on the 22 autosomes. The positions of the markers were from the Marshfield map (http://research.marshfieldclinic.org/genetics, Broman et al. 1998), using the sex-averaged positions converted to the Haldane mapping function. Phenotypic information is provided for 2885 persons (1213 from the original cohort). Detailed information about the Framingham Heart Study is given at http://www.nhlbi.nih.gov/about/framingham/index.html.

6.1.1.2 Condensation and trimming of pedigrees

The given pedigrees had to be condensed and trimmed to enable efficient multipoint linkage analysis with Merlin (Abecasis et al. 2002). Condensation was done without losing linkage information since only untyped individuals were discarded. Here, ungenotyped persons without children and untyped founders with only one child were removed, since they are not informative for linkage. After this step, four families were removed because they had no informative relationship left and four families fell into two unrelated branches. Finally, 14 families, which were still too large to allow some of the planned analyses, were trimmed by breaking some relationships that carried the least linkage information. This

resulted in a total of 346 pedigrees with 2656 individuals used in all analyses. The pedigree size ranged from four to eighteen individuals in two to four generations.

6.1.1.3 Phenotype definition

The phenotype height was investigated as the maximum of the available height measurements over the age of 18 years. Regression models for height were built for each sex in the original and the offspring cohort separately, adjusted for age at first examination to account for the different years of birth. The model for the i^{th} individual is:

$$\max(height)_i = \mu + \beta\, a_i + e_i$$

with μ - overall mean, a_i - age at first examination and e_i - residual. The standardized residuals are approximately normally distributed and were taken as height variables in the linkage analyses.

6.1.1.4 Linkage analysis methods

Multipoint linkage analyses for the height phenotype were done with the variance components (VC) models implemented in Merlin (Abecasis et al. 2002) and SOLAR (Almasy and Blangero 1998) and the inverse regression method (Sham et al. 2002) implemented in Merlin-Regress.

Variance components methods model the phenotypic variance which is explained by the estimated identity-by-descent sharing at a chromosomal position. Merlin (Regress and VC) calculates exact IBD sharing probabilities using the Lander-Green algorithm with sparse gene flow trees and can handle pedigrees up to about 20 individuals for multipoint analysis (Abecasis et al. 2002). On the other hand, SOLAR estimates multipoint IBD sharing probabilities with a generalization of the Fulker method (Fulker et al. 1995; Almasy and Blangero 1998) and has no restriction on the pedigree size.

6.1.2 Results

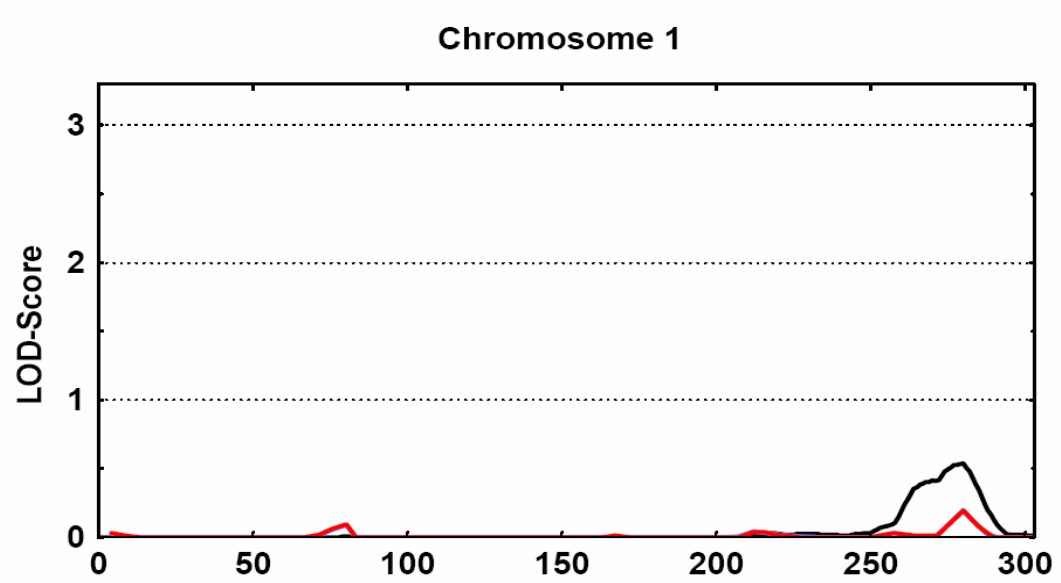
For chromosomes 6p, 6q, 9, 12, 14, 18 and 22, LOD scores greater than 1 for adult height were obtained with at least one analysis method. The strongest evidence for linkage to height was found near the q-ter of chromosome 6, with a

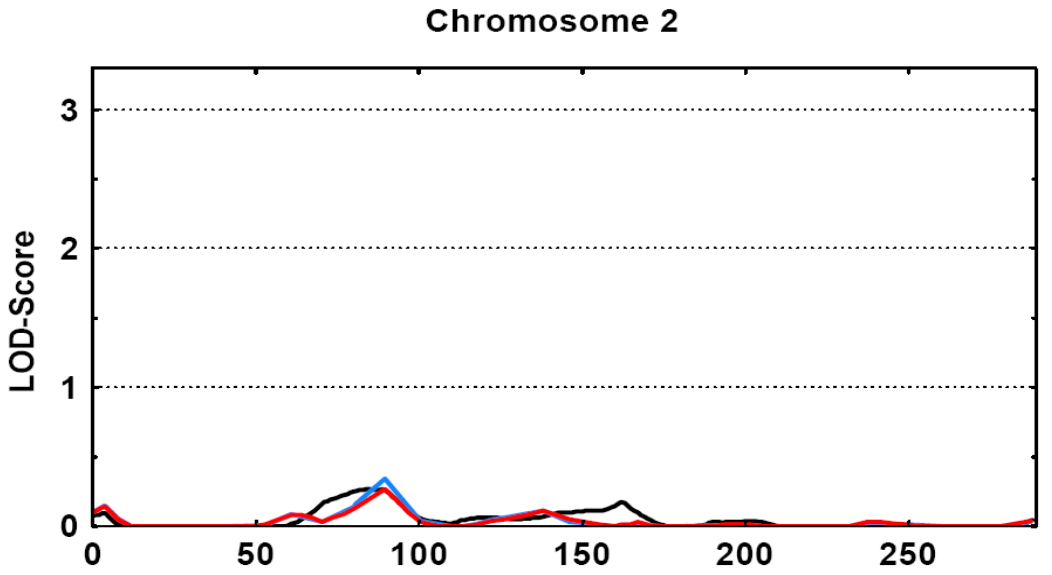
LOD score of 2.45 for Merlin-Regress and a 1-LOD support interval spanning from 190 to 204 cM. The variance components methods gave LOD scores of 1.83 and 1.67 at the same position (Figure 8). Table 21 gives all LOD scores greater than 1. The heritability of maximum height was estimated by the VC methods as 0.8.

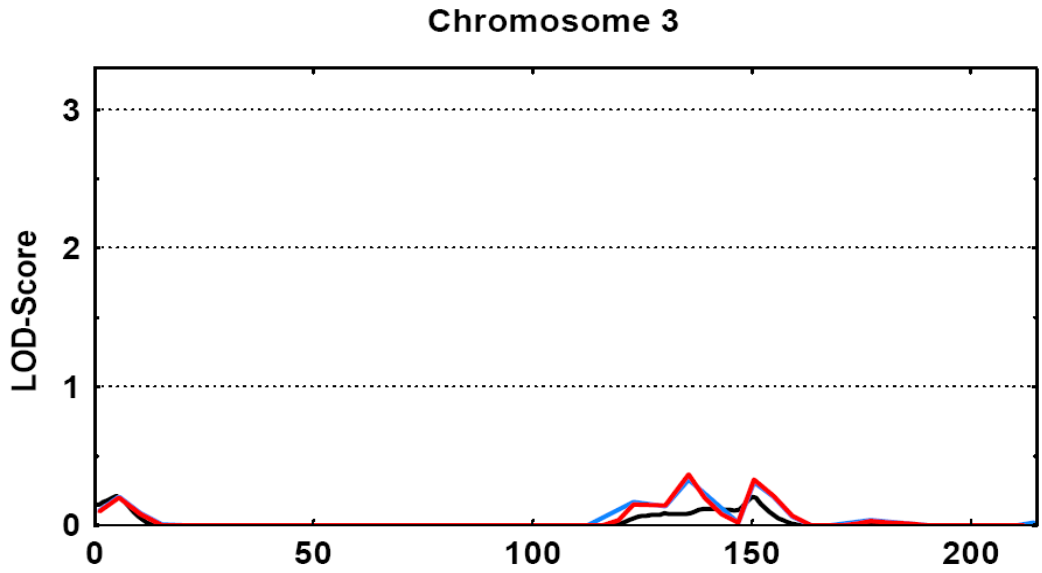
Table 17: Multipoint LOD Scores >1 for adult height with different methods.

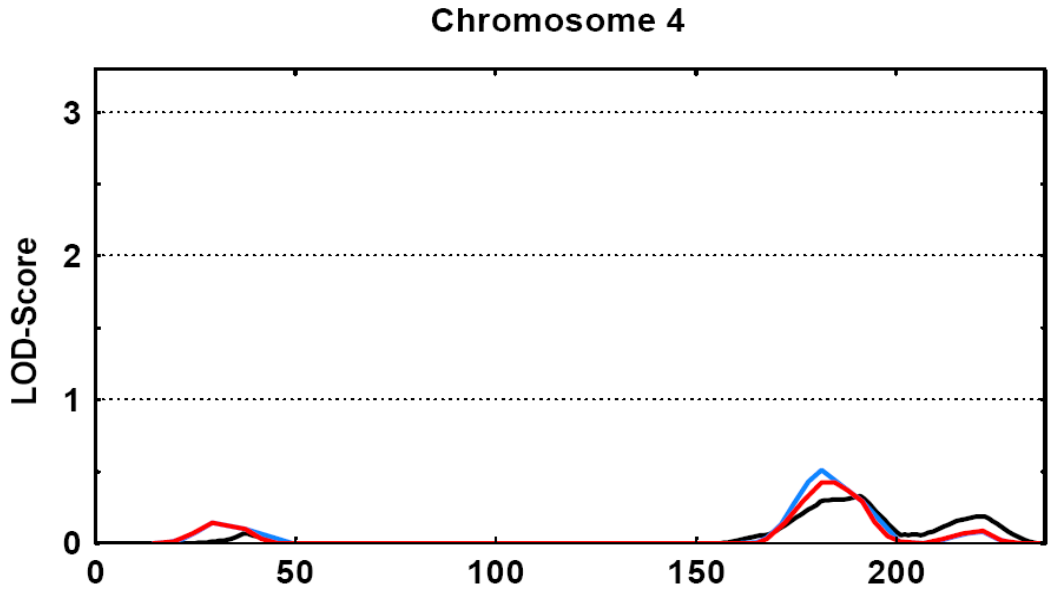
Chromosome	Marker	Peak Position (cM)	LOD Scores		
			SOLAR	Merlin-VC	Merlin-Regress
6	D6S2434	23-28	1.24	1.34	1.36
6	GATA184A08	159-162	1.19	1.06	1.14
6	D6S503	200-201	1.83	1.67	2.45
9	D9S319	60-61	1.22	1.15	1.56
12	D12S398	70-77	0.48	1.33	1.70
14	D14S742	5-11	1.27	0.80	1.02
14	D14S1426	137	1.35	1.49	1.58
18	D18S1364	111	1.54	1.44	1.73
22	D22S345	19-21	1.07	1.16	1.28

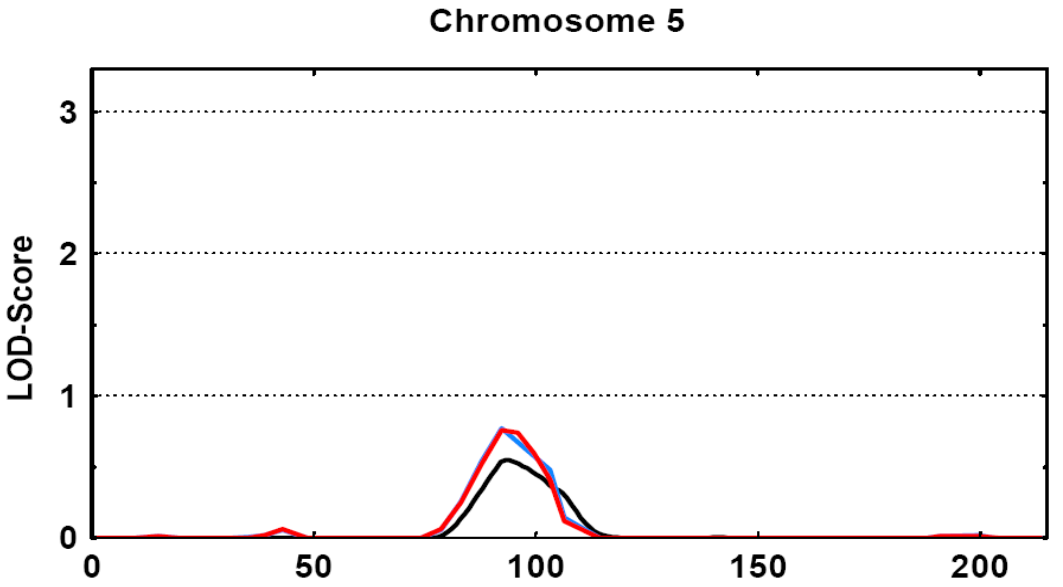
Figure 8: Multipoint LOD scores for height from SOLAR, Merlin-VC and Merlin-Regress for the 22 autosomes. Analysis method: SOLAR (black line), Merlin-VC (red line), Merlin-Regress (blue line). For ease of presentation, negative LOD scores from Merlin-Regress were set to 0. Positions are given in cM.

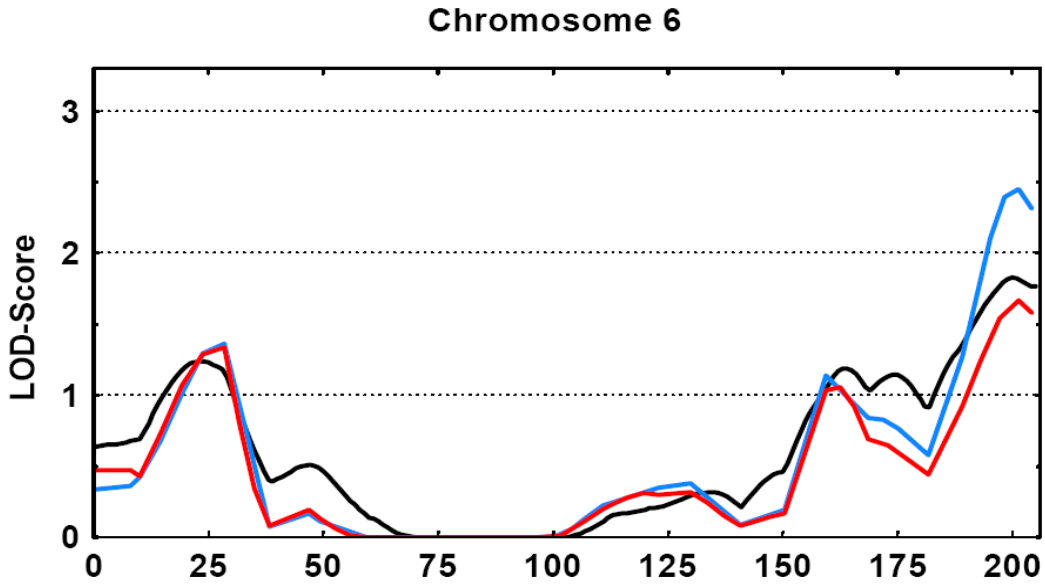


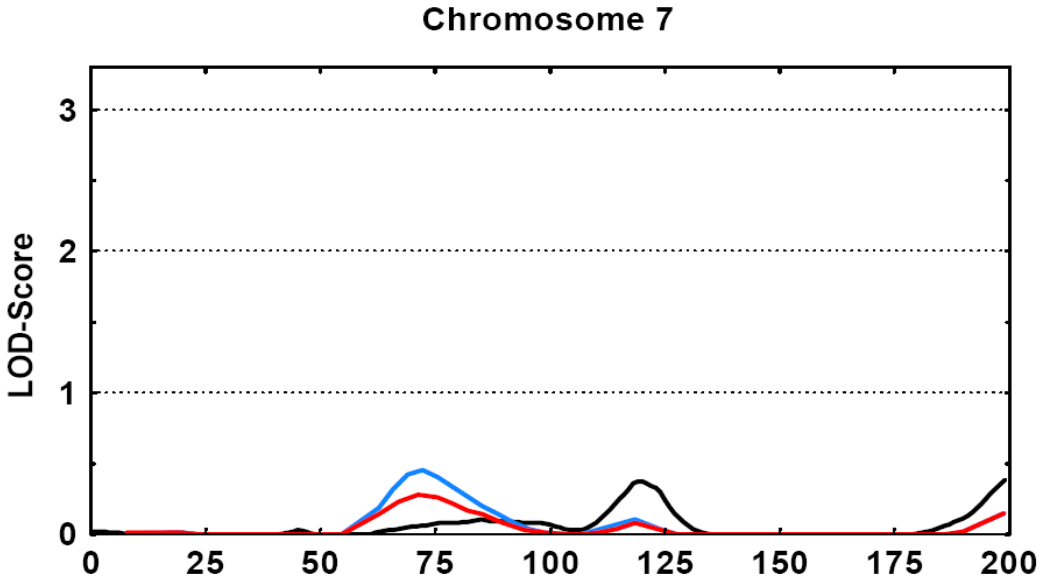

Chromosome 2
LOD-Score
3
2
1
0
0
50
100
150
200
250

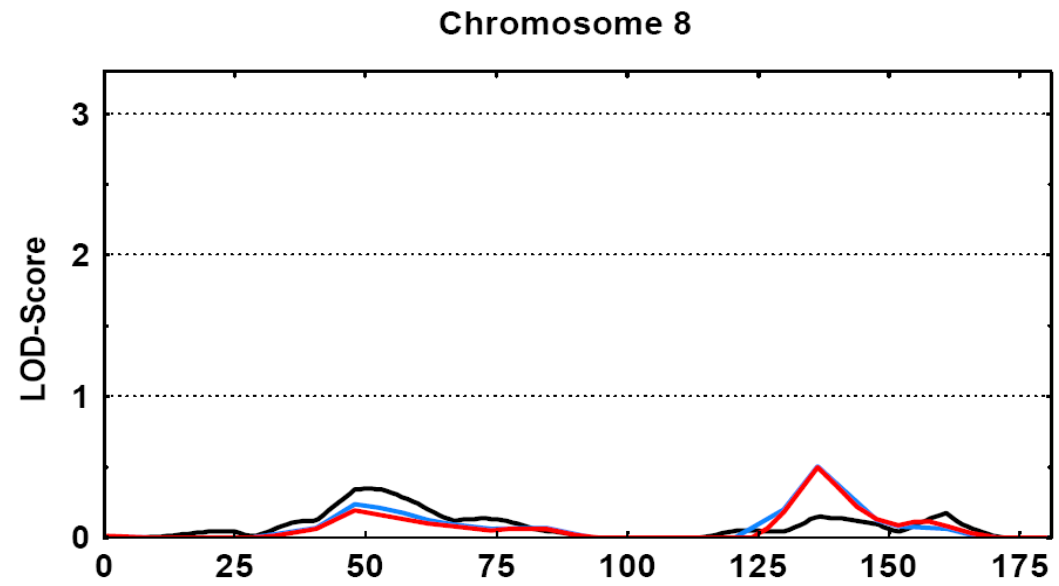

Chromosome 3
LOD-Score
3
2
1
0
0
50
100
150
200

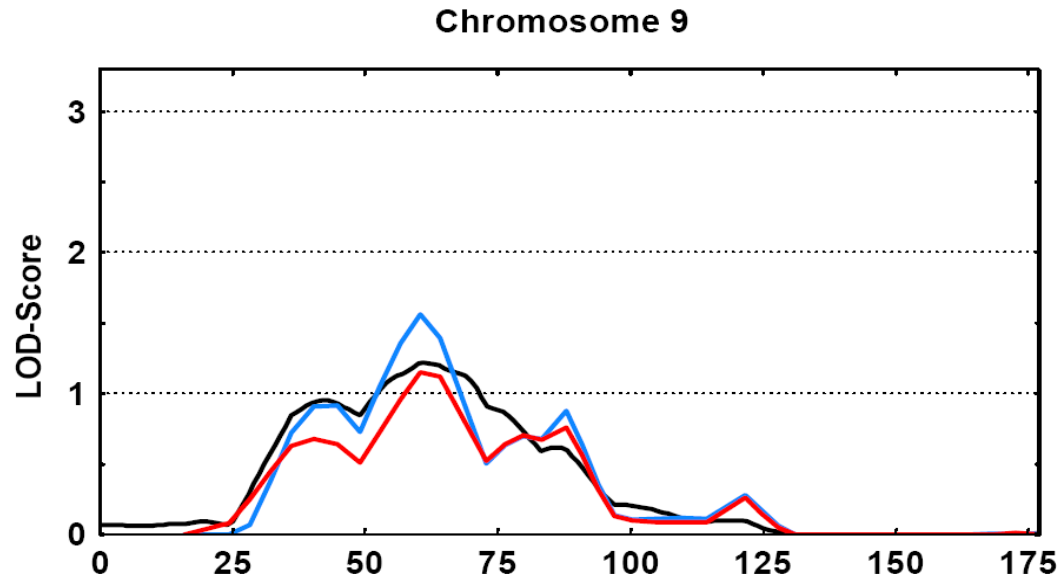

Chromosome 4
LOD-Score
3
2
1
0
0
50
100
150
200

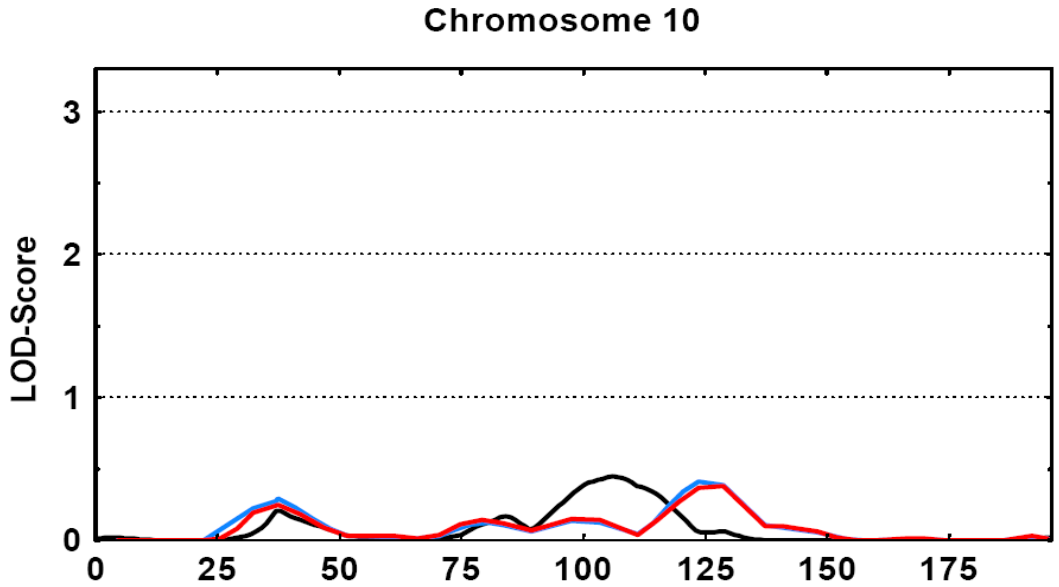

Chromosome 5
LOD-Score
0
1
2
3
0
50
100
150
200

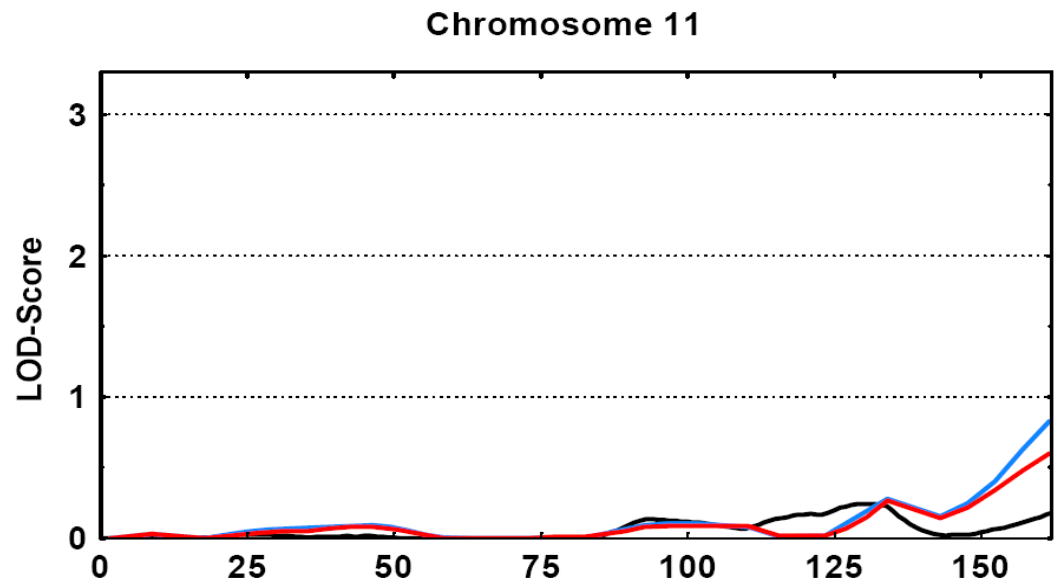

Chromosome 6
LOD-Score
0
1
2
3
0
25
50
75
100
125
150
175
200

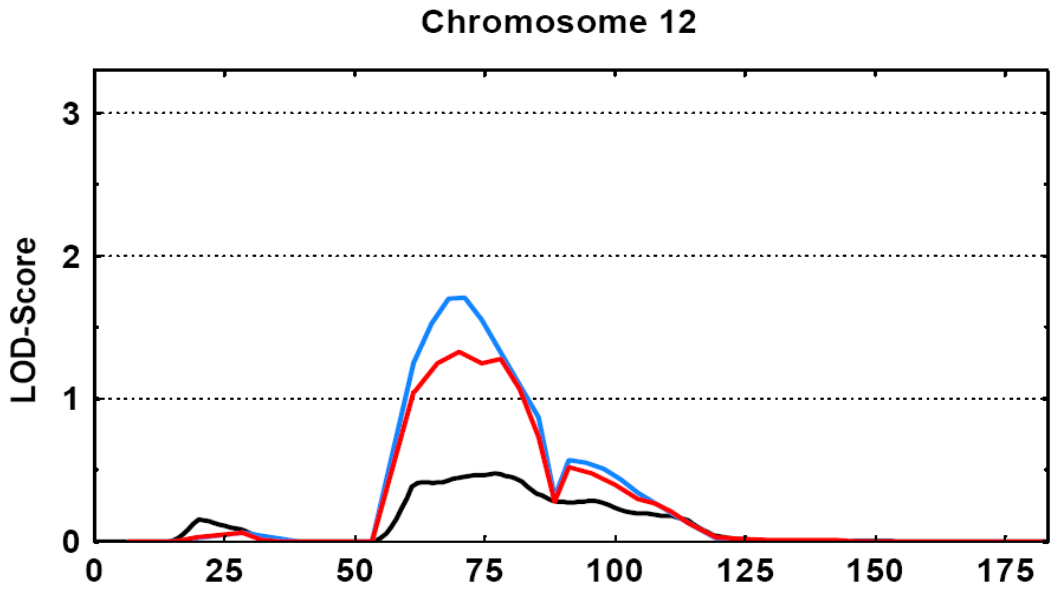

Chromosome 7
LOD-Score
0
1
2
3
0
25
50
75
100
125
150
175
200

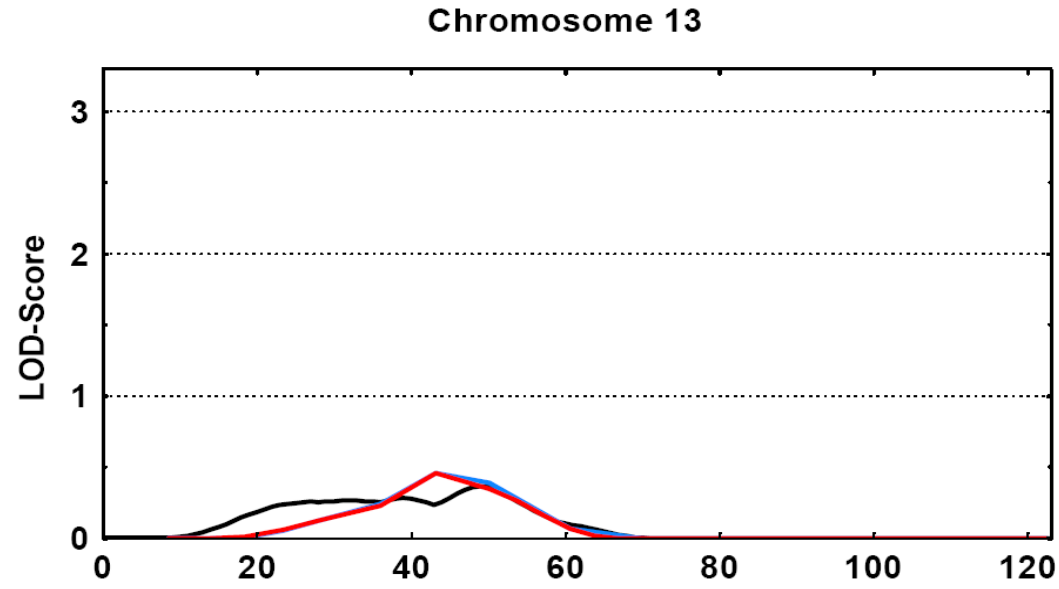

Chromosome 8
LOD-Score
3
2
1
0
0
25
50
75
100
125
150
175

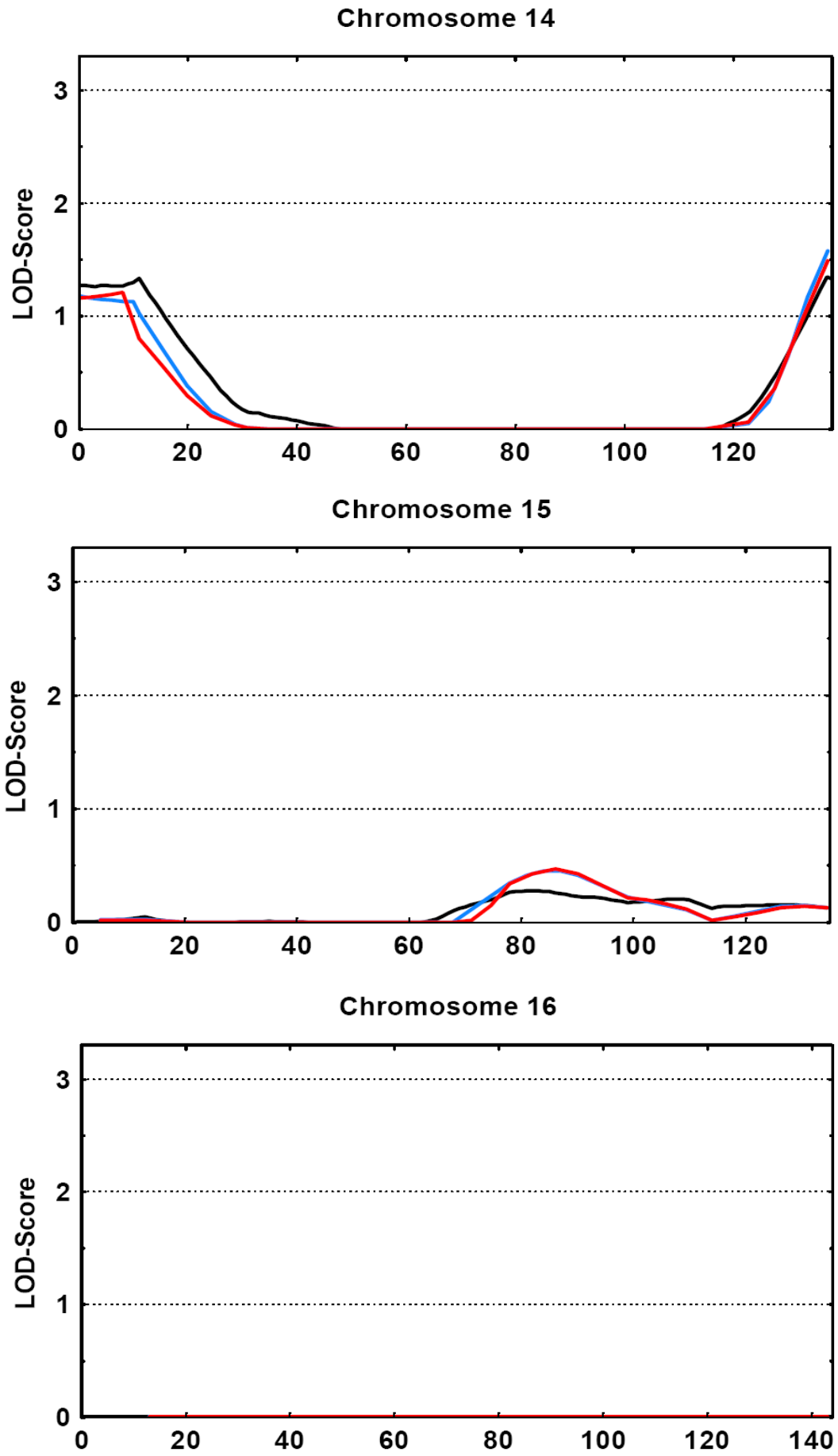

Chromosome 9
LOD-Score
3
2
1
0
0
25
50
75
100
125
150
175

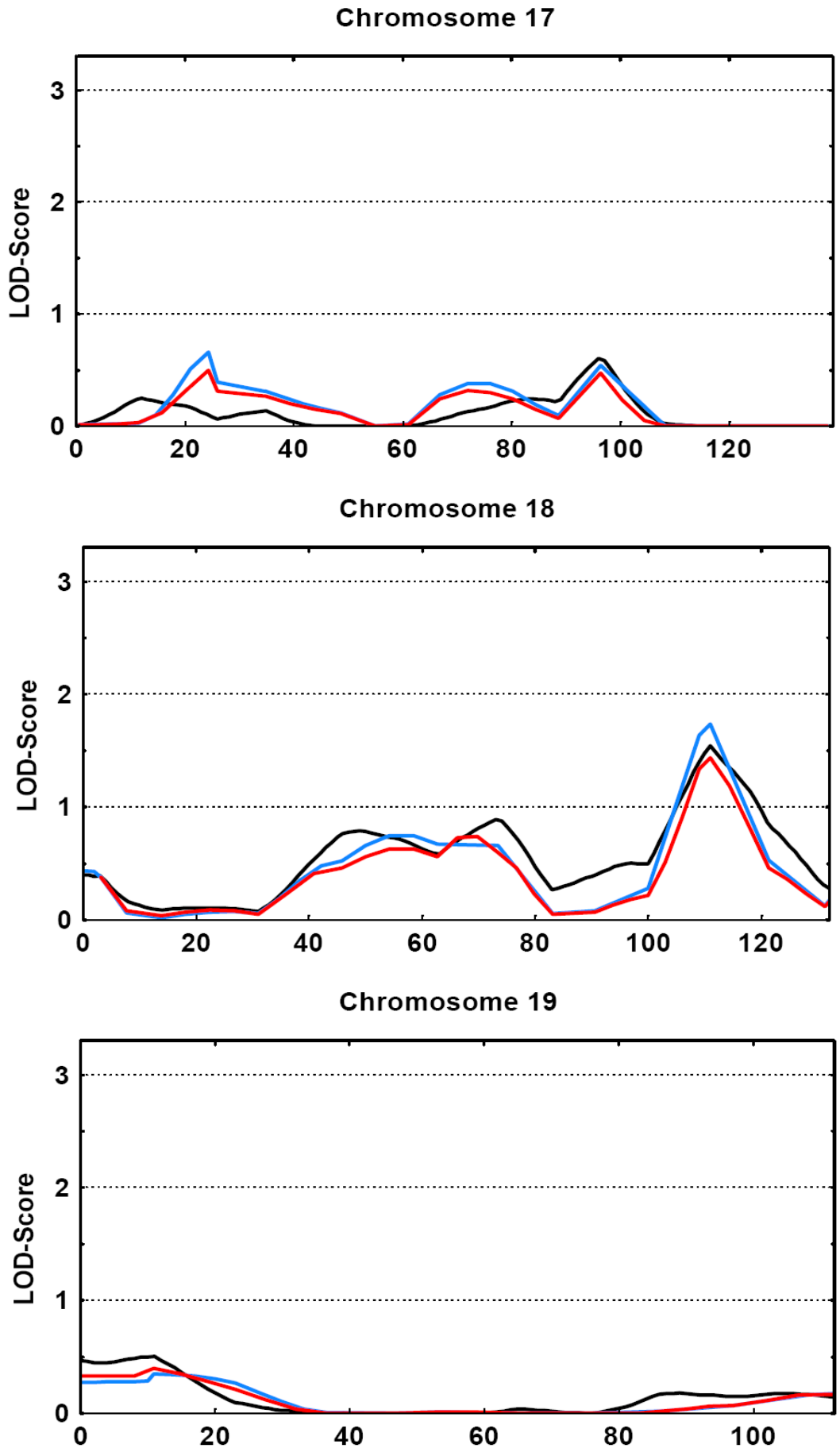

Chromosome 10
LOD-Score
3
2
1
0
0
25
50
75
100
125
150
175

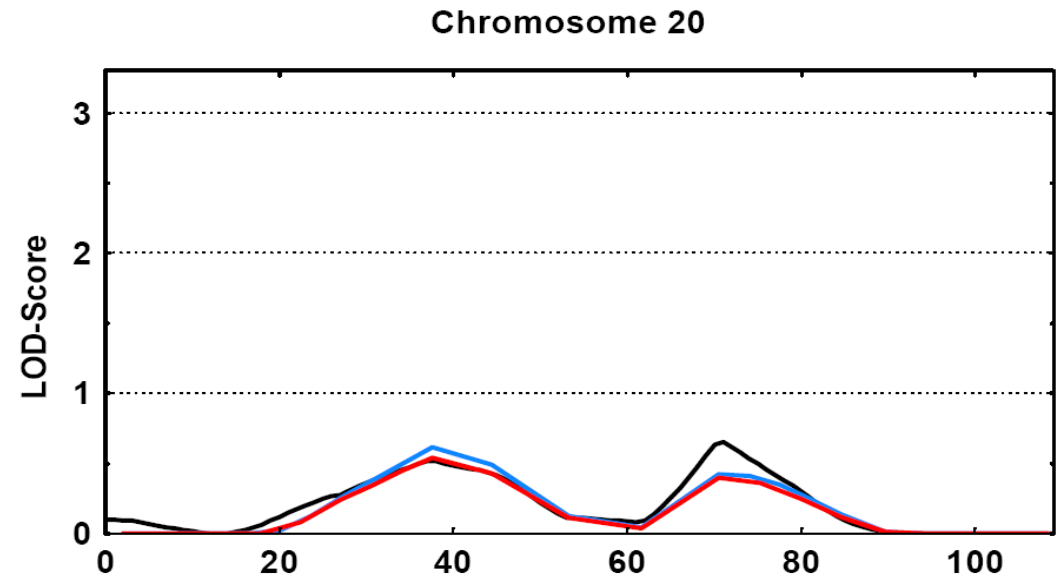

Chromosome 11
LOD-Score
0
1
2
3
0
25
50
75
100
125
150

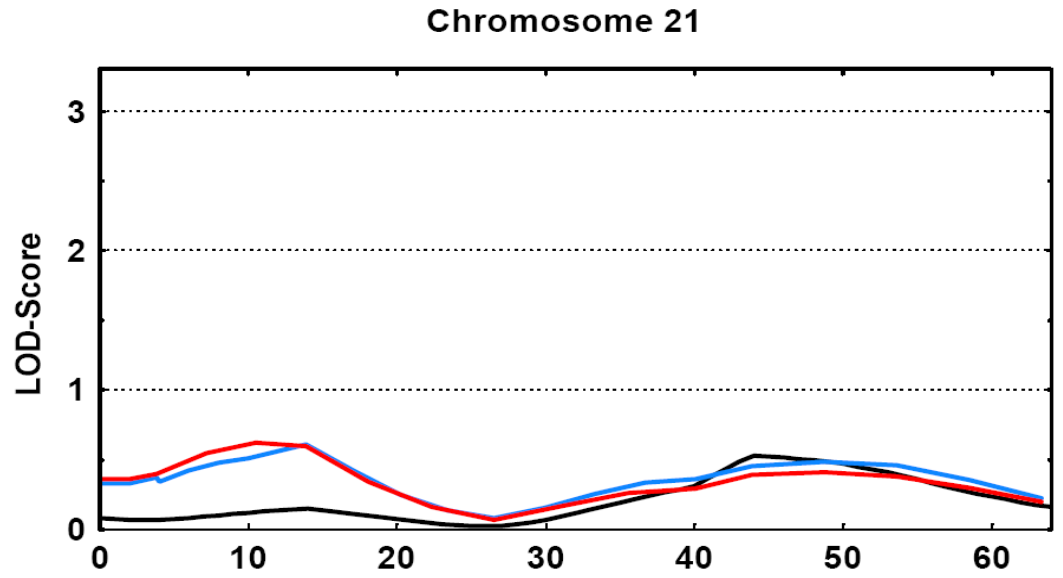

Chromosome 12
LOD-Score
0
1
2
3
0
25
50
75
100
125
150
175

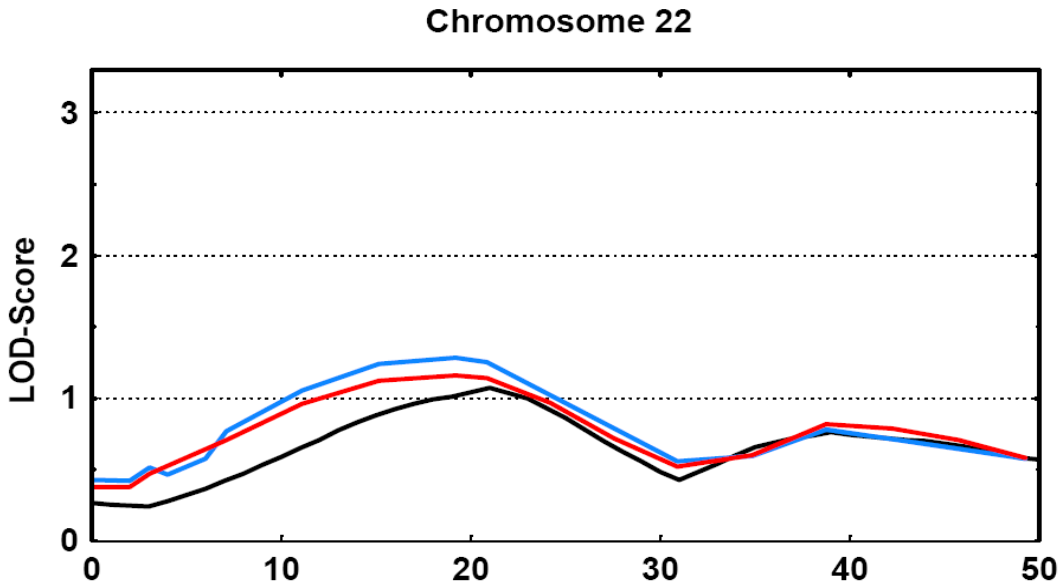

Chromosome 13
LOD-Score
0
1
2
3
0
20
40
60
80
100
120


Chromosome 14
LOD-Score
0
1
2
3
0
20
40
60
80
100
120
Chromosome 15
LOD-Score
0
1
2
3
0
20
40
60
80
100
120
Chromosome 16
LOD-Score
0
1
2
3
0
20
40
60
80
100
120
140


Chromosome 17
LOD-Score
0
1
2
3
0
20
40
60
80
100
120
Chromosome 18
LOD-Score
0
1
2
3
0
20
40
60
80
100
120
Chromosome 19
LOD-Score
0
1
2
3
0
20
40
60
80
100


Chromosome 20
LOD-Score
0
1
2
3
0
20
40
60
80
100


Chromosome 21
LOD-Score
0
1
2
3
0
10
20
30
40
50
60


Chromosome 22
LOD-Score
0
1
2
3
0
10
20
30
40
50

6.1.3 Discussion

A linkage genome scan for the quantitative phenotype adult height was performed in the extended pedigrees of the Framingham Heart Study. Analyses were conducted using two variance components approaches (SOLAR and Merlin-VC) and one regression method (Merlin-Regress), which gave very similar results. Even though the power of linkage analysis was substantially reduced since for many founders no DNA was available, this population-based and unselected sample has been a good example for the successful identification of linked regions. The results indicate that for moderately to highly heritable traits the analysis of phenotypically well characterized but unselected and rather large samples of extended pedigrees is promising. Other such large epidemiological cohort studies, where many covariables are carefully collected, can be valuable and efficient tools in studying the genes and interactions between genes and environmental factors in common complex diseases.

A first comparison of the linkage results in this study with previous genome scans for the same phenotype identified several regions of potentially overlapping linkage findings. In particular, on chromosome 6q there was a broad peak with a maximum LOD of 2.45 at 201 cM. Interestingly, Hirschhorn et al. (2001) and Xu et al. (2002) reported LODs of 3.85 at 159 cM and 3.06 at 155 cM, respectively. In this region a LOD score of 1.19 was obtained and it remains unclear if the maximum LOD scores on 6q result from the same locus. Substantial corroborative evidence exists also from Hirschhorn et al. (2001), Xu et al. (2002) and Perola et al. (2001) for the regions on chromosomes 6p, 9, 12, 14, 18 and 22 (see table 22).

Table 18: Multipoint LOD scores for adult height in different scans.

Chromosome and Region (cM)		Framhingham study	Xu et al. 2002	Hirschhorn et al. 2001	Perola et al. 2001
6p	0-30	1.36	<1	1.08	1.4
6q	155-200	2.45	3.06	3.85	0
9	42-66	1.53	2.09	2.01	0.15
12	56-80	1.55	1.86	3.35	0.82
14	11-47	1.02	<0.8	<1	1.67
18	72-116	1.73	<0.5	1.77	1.71
22	0-27	1.28	<0.8	1.95	0.54

Upper limits for the LOD scores <1 from Hirschhorn et al. (2001) and Xu et al. (2002) were estimated from their figures.

However, there was no overlap with the putative linkage regions reported by Thompson et al. (1995) and Wiltshire et al. (2002). Deng et al. (2002) reported a LOD score of about 1 on chromosome 18 at 75 cM. When comparing the results from these genome scans, differences between the studies have to be considered. Whilst all but one study investigated individuals of Caucasian origin (Thompson et al. (1995) analyzed Pima Indians), differences in sampling, sample size, pedigree structure and marker sets were more pronounced.

To exactly evaluate the combined evidence for linkage to the phenotype adult stature in these regions and throughout the genome, a meta-analysis of all published genome scans was therefore conducted.

6.2 Data sets for meta-analysis

For this meta-analysis, all published linkage genome scans up to August 2004 for the phenotype adult height were intended to be included (Deng et al. 2002; Geller et al. 2003; Hirschhorn et al. 2001; Perola et al. 2001; Thompson et al. 1995; Wiltshire et al. 2002; Wu et al. 2003; Xu et al. 2002). The 8 identified studies reporting linkage genome scans for adult height vary in several aspects of study design. An overview of important points is given in table 17. The first molecular genetic study of adult height was published in 1995, the others between 2001 and 2003. Most studies included Caucasian families, from different countries in Europe and North America, one used Pima Indians and one had subsamples of African American, Mexican American and Asian ethnicity. Five studies used a single sample each and 3 studies combined four to eight subsamples. The collection includes two cohort studies of families not ascertained for a specific phenotype and 20 studies that recruited for phenotypes such as hypertension, asthma or diabetes. Family structures that were recruited also differ substantially between studies from sib pairs to 3-4 generation pedigrees and average family sizes range from 2.4 to 11.9 individuals. A detailed description of each study is given below.

Table 19: General characteristics of the published genome scans for height.

First author	Year of Publication	Population, Country or Region	No of separate samples	Phenotype for which studies were ascertained
Thompson	1995	Pima Indians	1	Cohort study, no specific phenotype
Hirschhorn	2001	Finland, Sweden, Canada (Quebec)	4	Diabetes 2, coronary heart disease
Perola	2001	Finland	5	Hypertension, obesity, osteoarthritis, migraine, familial hyperlipidemia
Wiltshire	2002	UK	1	Diabetes 2
Xu	2002	Netherlands	1	Asthma
Deng	2002	US whites	1	Osteoporosis, obesity
Geller & Dempfle	2003	US whites	1	Cohort study, no specific phenotype
Wu	2003	European American, African American, Mexican American, Japan and China	8	Hypertension

6.2.1 Study identification and data collection

All published linkage genome scans up to August 2004 for the phenotype adult height were intended to be included in this meta-analysis. Segregation analyses and association studies or mutation screens of candidate genes are not included. Relevant studies were identified through searches of PubMed (http://www.ncbi.nlm.nih.gov) using different combinations of the search terms “genome”, “scan” or “screen”, “height” or “stature” and “linkage”. Abstracts of the PubMed search results were read and yielded six eligible publications (Perola et al. 2001; Hirschhorn et al. 2001; Wiltshire et al. 2002; Xu et al. 2002; Deng et al. 2002; Wu et al. 2003). Additionally, the identified papers and their reference lists were searched for other publications. This resulted in the identification of the paper by Thompson et al. (1995) where the title, abstract and key words do not reveal that a genome scan was done. The analysis of the Framingham Heart Study genome scan for the phenotype height was done in the context of the Genetic Analysis Workshop 13 (GAW13) and was published in December 2003 (Geller et al. 2003). The four studies published after August 2004 (Liu et al. 2004; Sale et al. 2005; Sammalisto et al. 2005; Willemsen et al. 2004) were not included. The study of Liu et al. (2004) was performed in an extended sample of that used by Deng et al. (2002).

For a meta-analysis of genome scans, it is necessary to have the results of each study at corresponding locations along the whole genome available to combine them appropriately. These could be results from common markers that were genotyped in several studies or at evenly distributed positions, e.g. every cM. Publications of genome scan studies usually do not include the results in such detail, but report only the markers or positions with the most interesting results exactly and include all other results in a figure. Therefore the authors of all published studies were contacted and asked for the detailed results, including the names and positions of all genotyped markers and LOD scores or other linkage statistics for all available positions. Details of the study design that were not sufficiently clear from the publications were also requested, such as information on family structures and numbers of genotyped and phenotyped individuals. The data used for the meta-analysis were only a more detailed version of the already published summary statistics (mostly LOD scores) and not any individual data such as genotypes.

Six of the contacted seven authors agreed to supply the necessary detailed results for the meta-analysis. The authors of Thompson et al. (1995) refused to make their results available, even after an intervention of the journal's editor. The American Journal of Medical Genetics where this study was published does not currently have a policy on authors' obligations to make data available to other interested scientists. Thompson et al. (1995) report in their publication only results on chromosome 20, where some markers showed significant p-values. No data for the other chromosomes are given. Therefore the results for chromosome 20 could not be included in the meta-analysis as this would bias the combined conclusions (publication bias).

6.2.2 Description of published genome scans

In the following, the design and important results of each of the separate genome scans are briefly presented. An overview of design and statistical methods of all scans is given in tables 18 and 19 for convenient comparison of important features across scans.

Table 20: Design features of the published genome scans for height (as reported in the individual studies).

First author	Individuals phenotyped and genotyped	Family structures	Average family size	No of families	No of genotyped markers
Thompson	924	Sib pairs with parents	4.1	226	160
Hirschhorn	2111	Sib pairs to extended pedigrees	4.8	483	Botnia: 920 Finland: 504 Saguenay: 362 Sweden: 504
Perola	580	Sib pairs	2.5	247	530
Wiltshire	1377	Sibships	2.4	573	418
Xu	962	2-3 Generations	4.8	200	366
Deng	630	Larger pedigrees	11.9	53	380
Geller & Dempfle	1702	2-3 Generations	7.7	346	401
Wu	6752	Sibships / nuclear families	2.7	2508	372

Table 21: Overview of statistical analyses in published genome scans for height (NR: not reported, VC: Variance components, SDS: Standard deviation scores).

First author	Transformation	Covariates	Separate models for	Estimated heritability of height	Linkage Statistic	Software used
Thompson	NR	NR	NR	NR	Haseman-Elston	SAGE
Hirschhorn	SDS	Age	Sex, study group	0.7 - 0.95	VC	Genehunter2
Perola	Square root, SDS	Age	Sex, study group	0.69	VC	SOLAR
Wiltshire	Square root, SDS	Age, diabetes status	Sex	0.89	VC	Genehunter2
Xu	SDS	Age	Sex	0.78	VC	SOLAR
Deng	-	Age	Sex	0.73	VC	SOLAR
Geller & Dempfle	SDS	Age	Sex, cohort	0.80	VC	SOLAR / Merlin
Wu	SDS	Age	Sex, study group	0.75 - 0.98	VC / Merlin-Regress	SOLAR / Merlin

As an overview of the important results of each scan, the different regions highlighted together with the obtained LOD scores is given in table 20. For a better comparison of these individual results and easier assessment of their concordance (or lack thereof), the LOD scores along the genome are plotted in figure 11 (together with the results of the meta-analysis), see below after the description of the individual scans (pages 98-105).

Table 22: Results of the individual scans as published. The two highest LOD scores and corresponding genetic region (approximate cM position according to the publication) are given. NR = not reported, '-' in the sample column denotes publications with just one sample.

First author	Sample	Highest LOD score	Region of highest LOD score		Second Highest LOD score	Region of second highest LOD score	
			Chromosome	cM		Chromosome	cM
Thompson	-	NR, p=0.0001	20	34-40	NR	-	-
Hirschhorn	Botnia	3.85	6	159	2.69	17	66
Hirschhorn	Finland	3.56	13	80	3.35	12	56
Hirschhorn	Sweden	3.40	7	150	1.95	22	0
Hirschhorn	Sanguenay-Lac-St.-Jean	1.35	17	40	1.26	4	13
Perola	combined	2.91	7	164	2.61	9	159
Wiltshire	-	3.17	3	9	2.26	7	103
Xu	-	3.06	6	150	2.09	9	50
Deng	-	2.14	5	144	1.95	X	10
Geller & Dempfle	-	1.83	6	200	1.54	18	111
Wu	European Americans	3.67	14	67	2.66	6	78
Wu	African Americans	2.25	1	136	NR	-	-
Wu	Mexican Americans	NR	-	-	NR	-	-
Wu	Asians	1.60	14	67	1.48	5	134

Thompson et al. scan (1995)

The Thompson et al. (1995) scan used a sample from a cohort study on the genetics of type 2 diabetes and obesity in Pima Indians from Arizona (USA), but included only non-diabetics in this analysis. Sib pairs and their parents from 226 families were genotyped for 160 markers and analyzed for the phenotype stature using the Haseman-Elston method (Haseman and Elston 1972). Three markers on

chromosome 20 (D20S66, D200S98 and D20S118; 34 to 40 cM) showed evidence for linkage to stature with nominal p-values of 0.0001 to 0.0003. D20S118 also showed evidence for linkage to leg length with a nominal p-value of 0.0002. A mutation screen of a candidate gene in this region, Bone morphogenetic protein 2 (*BMP2*) in 40 of the tallest and shortest individuals of the study group revealed one SNP, which was not associated with height in these 40 subjects. They concluded that other genetic variation in the region must be responsible for the observed linkage.

Hirschhorn et al. scan (2001)

Hirschhorn et al. (2001) used four different samples for their linkage analysis of height, a sample from the Botnia region in Finland, one from other parts of Finland and one from Southern Sweden, all with probands ascertained for type 2 diabetes, and one sample from the Saguenay-Lac-St.-Jean region in Canada with probands with either type 2 diabetes or coronary heart disease. The four samples were genotyped separately for different marker panels and the linkage analyses were also done separately using the variance components method implemented in Genehunter 2 (Pratt et al. 2000). This resulted in four genomic regions with LOD scores >3.3 on chromosomes 6, 7, 12 and 13 in different samples. In none of these regions, another sample reached LOD scores > 1 to support the respective result. On the other hand, there were another five regions (on chromosomes 4, 11, 17, 18 and 22) in which two samples each showed LOD scores between 1.35 and 2.69. A combined analysis of all four scans was done, using GSMA (Wise et al. 1999) which showed no genome-wide significant result, with the highest ranked bin overlapping the chromosome 6 region that was significant in the Botnia scan. This shows the difficulty in judging whether a region with one genome-wide significant result in a single scan and no replication of this in several other scans or rather e.g. two suggestive results in independent scans represent on the whole more statistical evidence for linkage. Only a quantitative meta-analysis, which incorporates important features of each scan such as sample size, can yield this.

Perola et al. scan (2001)

Perola et al. (2001) report a combined analysis of height in five samples from Finland, ascertained for osteoarthritis, familial combined hyperlipidemia, hypertension, obesity and migraine. Here, genotyping was done separately with only slightly different marker panels, but then the individual genotype and phenotype data were analyzed as one sample, again using variance components linkage analysis. Separate analyses of the samples were also done, but no detailed results were reported and could also not be obtained from the authors. The maximum LOD score on chromosome 7 is carried mostly by one sample and overlaps with the chromosome 7 peak in the Swedish sample by Hirschhorn et al. (2001).

Wiltshire et al. scan (2002)

Whiltshire et al. (2002) conducted a linkage analysis for stature in a single, large British/Irish sample that had been recruited for type 2 diabetes, including 573 sibships. The region of their highest LOD score on chromosome 3 is not highlighted in any of the other scans and their region on chromosome 7 is about 60 cM from the peaks on chromosome 7 obtained by Perola et al. (2001) and Hirschhorn et al. (2001). Indeed, at the Perola and Hirschhorn region (around 150-165 cM), Wiltshire at al. (2002) have a LOD score of 0.

Xu et al. scan (2002)

Xu et al. (2002) used a sample of 200 Dutch families ascertained through a proband with asthma to first perform a detailed segregation analysis of adult height followed by a genome wide linkage analysis. Heritability estimates were around 0.8 for most relative pairs used to calculate correlations, well in accordance with estimates reported elsewhere (Preece 1996; Silventoinen et al. 2000; Silventoinen 2003). There was also significant evidence of assortative mating for height, with a spouse correlation of 0.16. The segregation analysis gave the best fit for a mixed-recessive model, i.e. a model with a major recessive gene and residual polygenic effect, while a purely polygenic model fit the data significantly worse. The hypothetical recessive gene would explain 38% of the total variance of height in these families.

In the linkage analysis, the highest LOD (3.06) was obtained on chromosome 6, very close to the peak in the Botnia sample by Hirschhorn et al. (2001). Close to the second highest LOD (2.09 on chromosome 9) was another Botnia peak with a LOD of 2.01. The third highest peak (1.86 on chromosome 12) overlaps with yet another peak from the Hirschhorn et al. (2001) study in the Finland sample (LOD of 3.35).

Deng et al. scan (2002)

Deng et al. (2002) searched for linkage to adult height in a sample of white Americans of European descent which had been collected for a study on the genetics of osteoporosis. Each of the 53 extended pedigrees contained a proband with low bone mineral density (lowest 10 percentiles of population), which is correlated with height (r^2=0.12 in this sample, sex-adjusted), leading to significantly smaller heights in probands (sex- and age-adjusted difference of 8 cm). This ascertainment was accounted for in the variance components linkage analysis.

Deng et al. (2002) obtained their highest LOD score of 2.14 on chromosome 5 (close to a region with a LOD score of 1.48 in the Asian sample of Wu et al. (2003)), and a two-point LOD score of 1.95 on the X chromosome, which was not analyzed by the other studies and thus not included in this meta-analysis.

Wu et al. scan (2003)

Wu et al. (2003) analyzed the by far largest collection of more than 6700 individuals in more than 2500 families from eight different studies in four ethnic groups. These had all been recruited to study the genetics of blood pressure and hypertension. The genotyping was done for all eight samples together and with the same markers. The primary analysis as reported in the publication was done separately for the eight samples, but a combined analysis was also carried out. For the meta-analysis, the results of the separate analyses are used.

Two regions had elevated LOD scores in both European Americans and Asians (LODs of 3.67 and 1.6 on chromosome 14 and LODs of 2.26 and 1.48 on chromosome 5). On chromosome 6, a LOD of 2.66 was reached in the European American sample, but this is about 70 cM away from the LOD score peaks of Xu et al. (2002) and Hirschhorn et al. (2001, Botnia sample) on this chromosome.

The strongest support for linkage in the combined analysis was found on chromosome 7 (LOD 2.46), which did not stand out in any of the eight individual analyses, but is in the same region as the chromosome 7 peaks of Hirschhorn et al. (2001, Swedish sample) and Perola et al. (2001).

6.3 Meta-analysis methods

6.3.1 Available Data

From the published studies, VC LOD scores (from SOLAR or Genehunter) along the 22 autosomes were available, either at regular intervals, e.g. every 1 cM or at the genotyped markers and at certain positions between them. Additionally, the names and assumed positions of the genotyped markers were provided. Only the study by Deng et al. (2002) reported results for the X-chromosome, all other studies did not perform linkage analysis for this chromosome and therefore it was not included in the meta-analysis. For all studies, the sample size (usually number of phenotyped and number of genotyped subjects) and in more or less detail, the distribution of the sample in families (number of families, family structure such as sib pairs with or without parents) was given. For only two studies, additional data was obtained such as the Genehunter information content at the same positions as the LOD scores, but for almost all of the studies this was not available and thus could not be used in this meta-analysis.

The VC LOD score, which is the difference in $\log_{10}$ likelihoods between the restricted and unrestricted variance components models is equivalent to a regular parametric LOD score, i.e. twice the difference in $\log_e$ likelihoods between the restricted and unrestricted models yields a test statistic that is asymptotically distributed as a ½:½ mixture of a χ^2 variable with one degree of freedom and a point mass at zero (Self and Liang 1987; Almasy and Blangero 1998). Thus VC LODs can easily be converted to p-values or standard normal Z scores (and vice versa), which can then be used in a meta-analysis.

The VC LOD score is a measure of significance and not an effect estimate. Relevant effect size measures for variance components linkage analysis would be the (additive) genetic variance component or equivalently the locus-specific heritability (h^2). With these and the respective variance estimates, a fixed or random effects meta-analysis would be possible. Most genetic analysis programs (specifically Genehunter (Kruglyak et al. 1996) and SOLAR (Almasy and

Blangero 1998)) do not routinely provide these estimates, and they were not available for this meta-analysis. The program Merlin-Regress (Abecasis et al. 2002) on the other hand, which implements a regression approach to quantitative trait linkage analysis (Sham et al. 2002), normally provides estimates of the locus-specific heritability and its standard error in the detailed output.

6.3.2 Common marker map

In a meta-analysis of linkage studies, it is essential that linkage information of the same genetic location is combined over studies. This can easily be achieved if the same markers are analyzed in all studies, but becomes more difficult if the markers differ between studies. Even for markers that are common to more than one study, the assumed positions of markers for the analysis often differs, either through use of different marker maps, such as the Marshfield (http://research.marshfieldclinic.org/genetics/, Broman et al. 1998) and deCode maps (Kong et al. 2002), or because the position of the most proximal genotyped marker of each chromosome is arbitrarily designated as 0 cM. In this case, distances between markers may be equal across studies, but absolute positions in cM are not. Therefore a locus designated at the same cM position in different scans does not necessarily specify the same genetic location. This has to be corrected by constructing a consensus map, which includes unique cM positions for all markers genotyped in any study and subsequently repositioning all results (LOD scores) to this map. For the height meta-analysis, the Marshfield map (Broman et al. 1998) was chosen as the basis of this consensus map, because many of the individual studies had already used this map in their analysis (in several cases with the first genotyped marker on each chromosome given the position 0 cM). Of the total 1558 different markers in all studies, 1467 were included in the Marshfield map; only 91 are missing from the Marshfield map. The marker order was consistent with that used in the individual studies. Markers that were not included in the Marshfield map were placed according to their relative position assumed in the single studies, i.e. between the same flanking markers and at the same relative distance to them. The reported LOD scores were now designated new positions according to the positions of the corresponding markers on the consensus map. LOD scores from different studies pertaining to

the same marker are therefore assigned the same position and will then be combined in the meta-analysis.

Since the positions of the first and last marker on each chromosome differ quite a lot between studies, results of each study were only included 2 cM beyond the first and last genotyped marker.

6.3.3 Combination of linkage results

The only data available from all studies were variance components LOD scores, which are measures of significance (essentially a function of the likelihood ratio test statistic and thus equivalent to p-values) and no measures of effect size and corresponding standard errors could be obtained. Therefore only a meta-analysis method that uses p-values could be applied such as Fisher's p-value combination method. As Elston (1991) observes, Fisher's method is asymptotically optimal among essentially all methods to combine significance levels of independent tests. VC LOD scores were first transformed to p-values (via the χ^2 distribution), these were combined and the resulting p-value was back-transformed to a LOD score. This transformation to LOD scores is mainly done for ease of comparison with the results of the single studies. A slight modification is necessary since VC methods do not give negative LOD scores, even if allele sharing between affecteds is less than expected, but rather truncate the LOD score space at zero. So LOD scores of 0 are biased and should therefore not be assigned a p-value of 0.5 (as would be the case in parametric linkage analysis). Province (2001) pointed this out and proposed to give such LOD scores a p-value of $\frac{1}{2\log_{10}(2)} \approx 0.72$ in a meta-analysis as a way to avoid a bias in the combined result. This was used in the current meta-analysis.

There is no clear consensus in the literature, whether Fisher's method sufficiently accounts for different sample sizes of the combined studies (see Hedges and Olkin 1985, p. 38). The p-value results both from the observed effect size and the sample size. Thus, for a given effect size, the p-value will reflect the sample size, but since effect sizes are not expected to be constant across studies (not exactly the same IBD sharing at all locations in different samples), the p-value is no longer proportional to sample size. Weights, which could reflect different sample sizes, can be incorporated in Fisher's method by exponentiating

the p-values with the respective study weights (Good 1955). However, the distribution of the weighted test statistic is no longer a simple χ^2 distribution and is only known if the weights are distinct. Additionally, an optimal choice of weights has only been evaluated for special designs (Hedges and Olkin 1985).

The inverse normal method (Stouffer et al. 1949) on the other hand can easily be used unweighted or weighted with the relevant sample size for each study. For the weighted form, p-values (derived from LOD scores with Province's bias correction) were transformed to corresponding standard normally distributed Z-scores and these were weighted by study size and combined to a standardized Z-score which was back-transformed to a LOD score. These are defined as follows: let k be the number of studies, and n_j (j=1, …, k) be the relevant sample size of study j (as defined in the next section). Let, $p_j(t)$ (j=1, …, k) be the p-value of study j at position t (in cM) and $Z_j(t) = \Phi^{-1}[p_j(t)]$ be the standard normal quantile of $p_j(t)$. For the inverse normal methods, the combined scores are

$$Z_{unweighted}(t) = \frac{1}{\sqrt{k}} \sum_{j=1}^{k} Z_j(t)$$

$$Z_{weighted}(t) = \sum_j \frac{\sqrt{n_j} Z_j(t)}{\sqrt{\sum_j n_j}} \quad .$$

For comparison, results of three different meta-analyses are therefore presented, using Fisher's method, the unweighted inverse normal method and a weighted inverse normal method.

6.3.4 Relevant sample size

Weighting of the linkage results should be proportional to the sample size. In contrast to other types of studies, where each individual contributes the same amount of information and therefore the relevant sample size is just the number of individuals, this is not necessarily the case in a linkage study. The basic unit of information for linkage would be one informative meiosis (where a recombination or non-recombination is counted), so only non-founders in a pedigree can contain linkage information and while founders do not contain linkage information themselves, they can (through determination of phase) enhance the information of their children. Depending on the linkage test statistic used, an individual in a specific family constellation may carry different amounts of information, e.g.

unaffected offspring can also be counted as recombinant or non-recombinant in parametric linkage analysis (depending on the underlying penetrance model) but do not usually contribute to an affected sib-pair analysis (if parents are not typed, unaffected siblings may be used to improve IBD estimation). If the disease model is unknown, it is not known which meioses (dependent on phenotypes, i.e. disease genotypes) are informative, therefore it is reasonable to use e.g. the number of independent pairs (for a test statistic that is based on relative pairs) as a weighting factor (Sham 1998). For linkage analysis of quantitative traits, such as height, the information of an individual depends also on the phenotypic value. For variance components analysis, all genotyped and phenotyped non-founders can contribute information, provided there is at least one other genotyped and phenotyped family member besides his/her parents. For all included studies, the number of genotyped and phenotyped individuals was available (table 23). This was not presented separately for founders and non-founder, and further details regarding family structures could not be obtained from most authors. Therefore, the weighting was performed by using the number of genotyped and phenotyped individuals in each study (regardless of founder status).

If more detailed information on the family structures had been available, this could have been incorporated in an approach to calculating the relevant sample size. Tang and Siegmund (2001) showed that for variance components analysis of quantitative, oligogenic traits in nuclear families, the power of a sibship of size s is approximately the same as of $s(s-1)/2$ independent sib pairs. This justifies weighting k sibships of size s equally to $ks(s-1)/2$ independent sib pairs from different families. Tang and Siegmund (2001) also showed that in nuclear families where parents are also phenotyped and genotyped, the power for variance components linkage analysis is higher than without parents. Depending on the sibship size this ranges from about 15 % for sib pairs to 7 % for five sibs. So for nuclear families with or without parents, an “effective sample size” for VC linkage analysis could be calculated based on the approximately equal number of independent sib pairs. If family structures do not vary much within a study (as is usually the case) this can be used based on just the average family size in a study (and not the distribution of family sizes which is rarely provided). This has the advantage that it can often be calculated with just the published summary description of the sample.

Table 23: Family structures and sizes for linkage studies of height.

First author	Sample	No of individuals genotyped and phenotyped	No of extended pedigrees	No of nuclear families	Founders geno- & phenotyped	Mean family size	Range family size
Hirschhorn	Botnia	379	47	11	Partly	7	2-18
Hirschhorn	Finland	702	80	99	Partly	4.2	2-13
Hirschhorn	Sweden	683	89	94	Partly	4.1	2-11
Hirschhorn	Sanguenay-Lac-St.-Jean	347	2	61	Partly	6.7	3-15
Perola	combined	580	0	247	No	2.5	NR
Wiltshire	-	1377	0	573	No	2.4	NR
Xu	-	962	34	166	Yes	4.8	NR
Deng	-	630	53	0	Yes	11.9	3-99
Geller & Dempfle	-	1702	346	0	Partly	4.9	4-18
Wu	GenNet European Americans	598	0	NR[1)]	Yes	NR	NR
Wu	GenNet African Americans	595	0	NR[1)]	Yes	NR	NR
Wu	GENOA European Americans	749	0	NR[1)]	No	NR	NR
Wu	GENOA African Americans	611	0	NR[1)]	No	NR	NR
Wu	GENOA Mexican Americans	778	0	NR[1)]	No	NR	NR
Wu	HyperGEN European Americans	1100	0	NR[1)]	No	NR	NR
Wu	HyperGEN African Americans	1252	0	NR[1)]	No	NR	NR
Wu	Asians	1069	0	NR[1)]	No	NR	NR

[1)]it was only reported that the total number of families in all 8 data sets was 2508, and the average sibship size was 2.8 across all 8 data sets.

Other approaches to defining an appropriate effective sample size in linkage studies have also been described, especially in the context of meta-analysis. For meta-analyses of Bipolar Disorder and Schizophrenia (Levinson et al. 2002), the square root of the number of genotyped cases in each sample was used as the weighting factor for the respective study. For accompanying power analysis, they simulated samples of families which contained one ASP each (which is computationally easier and allows for easier interpretation of genetic model parameters). The number of ASP families was chosen to represent approximately equal linkage information to the samples of pedigrees of different size and structure as the original samples. Based on detailed information from one Schizophrenia linkage study, they estimated the equivalent number of independent ASPs as 1.39 times the number of genotyped cases minus the number of pedigrees. Such an approach is only reasonable if the pedigree structures and sizes are known in some detail and ideally, the appropriate conversion factor (here 1.39) should be determined for each study. They did not investigate whether the weighting of studies would be more powerful by using the equivalent number of independent ASPs as the effective sample size instead of using the number of genotyped cases (as they did for the actual meta-analysis).

6.3.5 Ethnicity and heterogeneity

Genetic heterogeneity between samples might lead to inconsistent results across studies. One important reason for heterogeneity could be the presence of causal alleles at very different frequencies in different populations, which would lead to linkage signals only in some populations but not in others. Ethnically homogeneous samples would reduce the risk of such genetic heterogeneity. Most of the samples analyzed for linkage to adult stature were of Caucasian origin, only the sample from the NHLBI Family Blood Pressure Program (Wu et al. 2003) includes three non-Caucasian sub-samples (African American, Asian and Mexican American), this supports the expectation that at least some common genetic effects can be identified in this sample. A limitation of meta-analysis methods that use only p-values is that no formal test for heterogeneity can be performed as there is no sensible way to obtain effect sizes from each study. The VC LOD scores available for this meta-analysis are equivalent to p-values in this respect.

6.4 Results

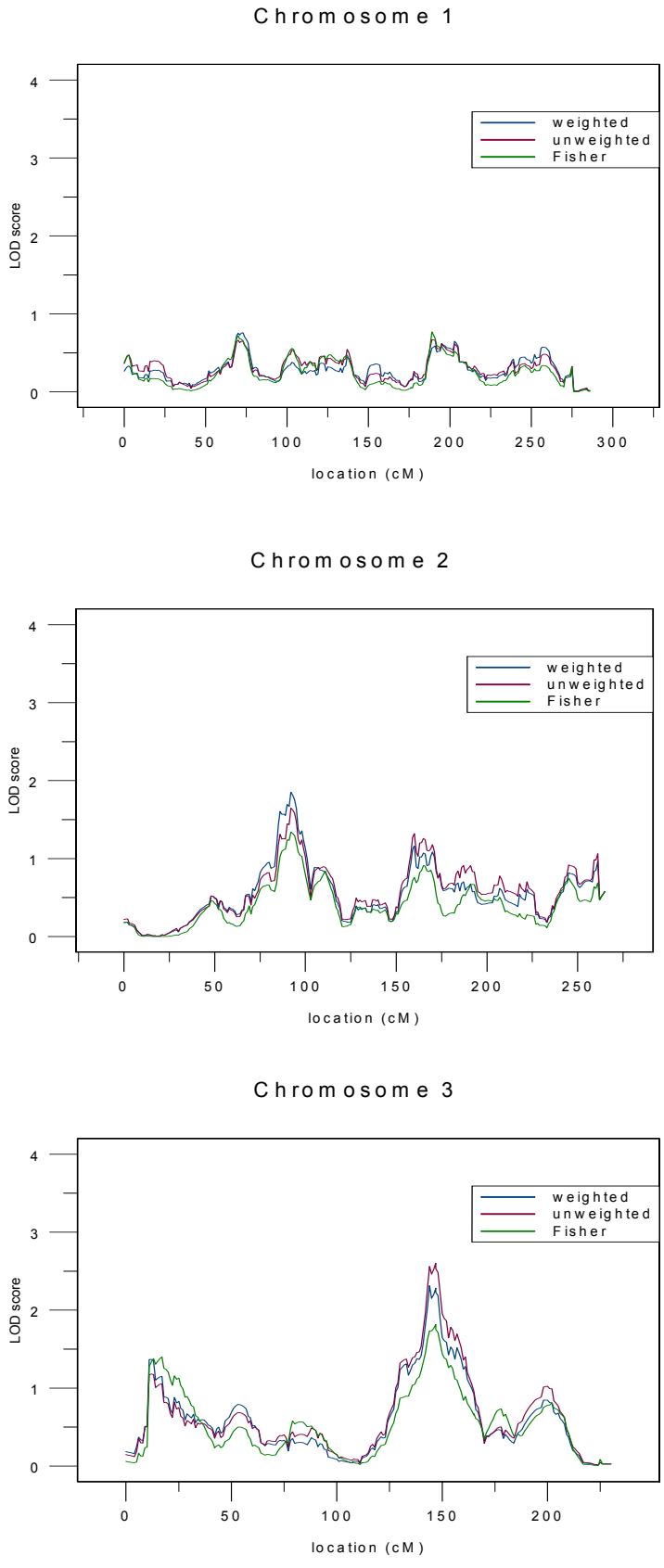
6.4.1 Comparison of Fisher's method and weighted and unweighted inverse normal meta-analysis

Figure 9 shows the results of the three applied meta-analysis methods on chromosomes 1 to 22: Fisher's method and the weighted and unweighted inverse normal meta-analyses. Table 24 gives the LOD scores of the different meta-analysis methods on selected chromosomes. In this comparison, it has to be kept in mind that on such a real data set, no general conclusions regarding power or type 1 error can be drawn, since the complex genetic architecture of human stature is currently not known in detail. Thus no genes are definitely known to be involved in the genetic regulation of height, nor can any genomic region be considered as not containing any genes of relevant effect on height.

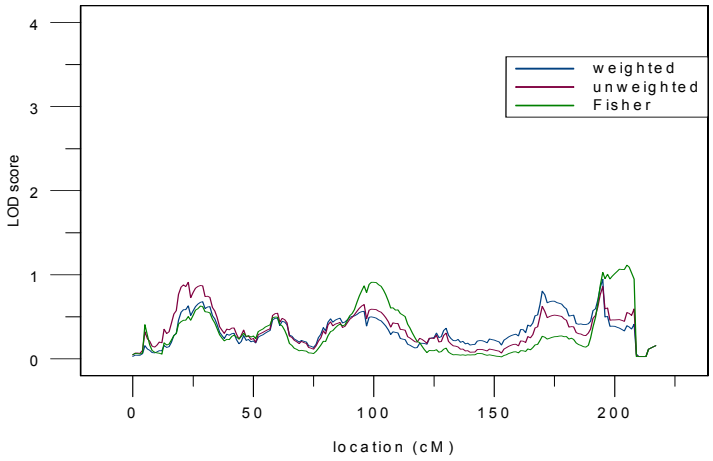
Table 24: LOD scores of different meta-analysis methods on selected chromosomes

		Maximum LOD in the region		
Chromosome	**Approximate position of maximum LOD (cM)**	**Fisher's method**	**Unweighted inverse normal**	**Weighted inverse normal**
3	144-147	1.82	2.6	2.31
5	130-138	1.92	1.25	1.27
6	157	4.44	3.4	3.33
7	175	3.63	4.59	4.13
9	62	3.95	4.26	4.08
12	64-65	4.12	3.95	4.08
13	98-99	2	1.71	1.34
14	63-66	2.93	1.73	1.41
15	98-101	2.01	2.52	2.66
17	76-77	2.24	2.26	2.0
20	29-33 and 62	2.48 (at 62 cM)	2.91	2.81

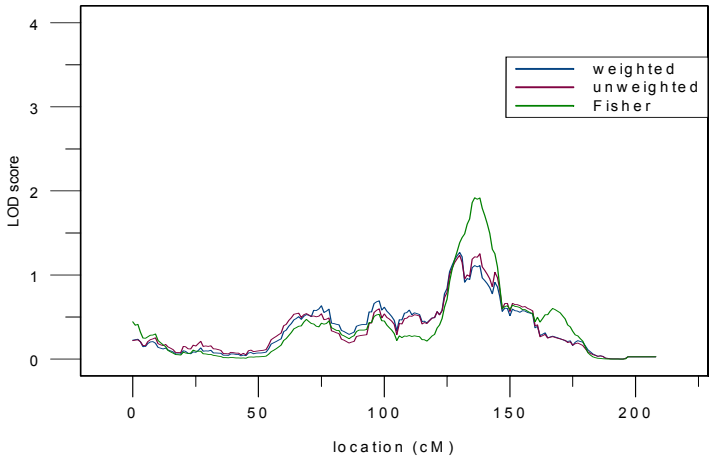
Figure 9: LOD scores obtained by the three different meta-analysis methods on chromosomes 1-22, the weighted and unweighted Z-score combination method and Fisher's method.



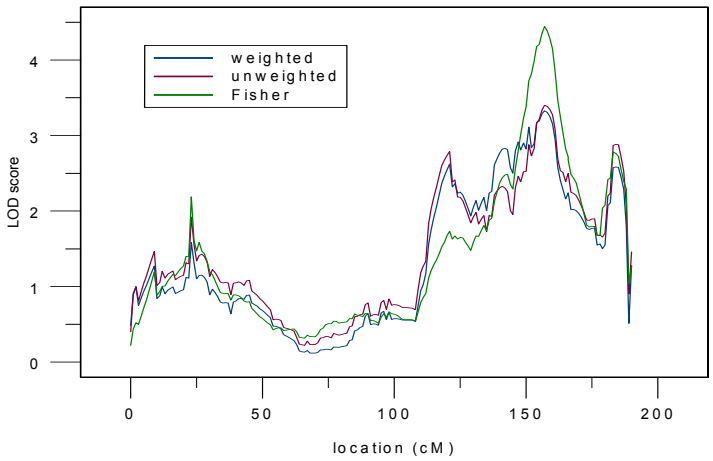
Chromosome 4



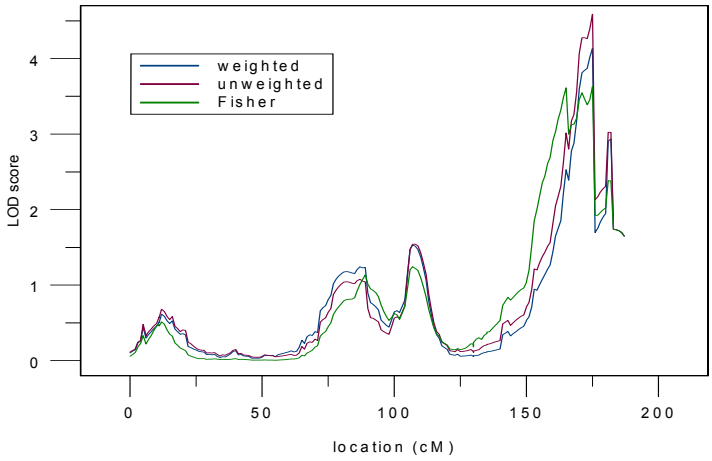
Chromosome 5



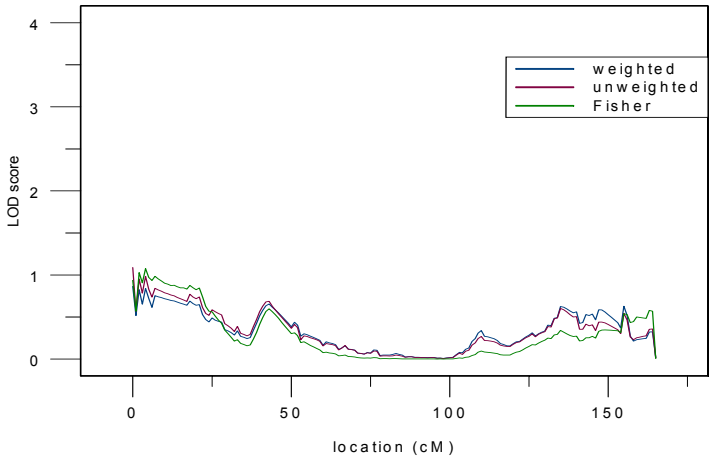
Chromosome 6



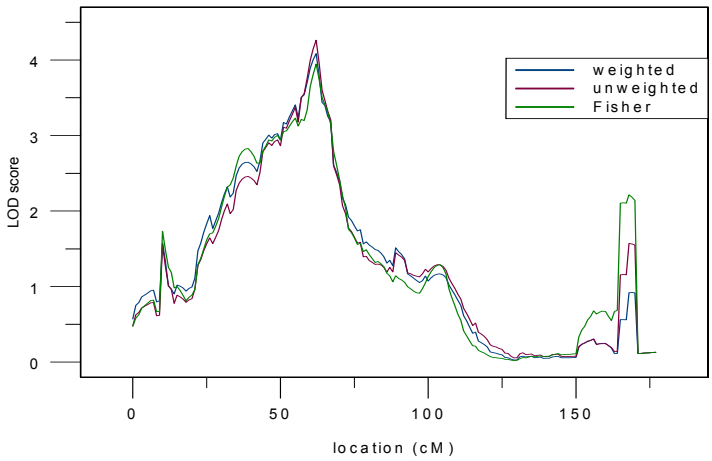
Chromosome 7



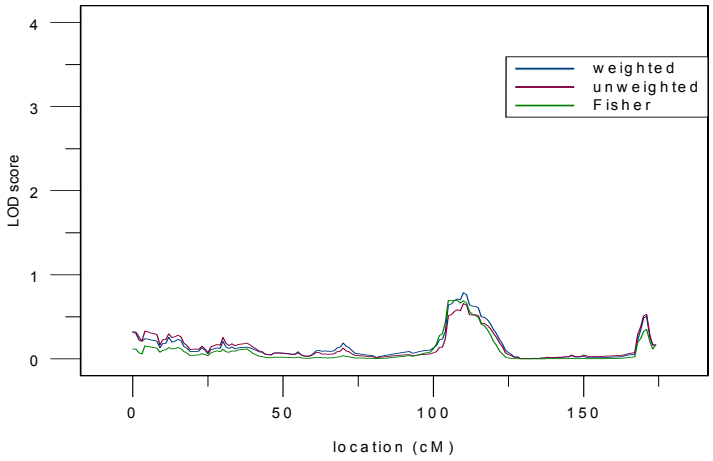
Chromosome 8



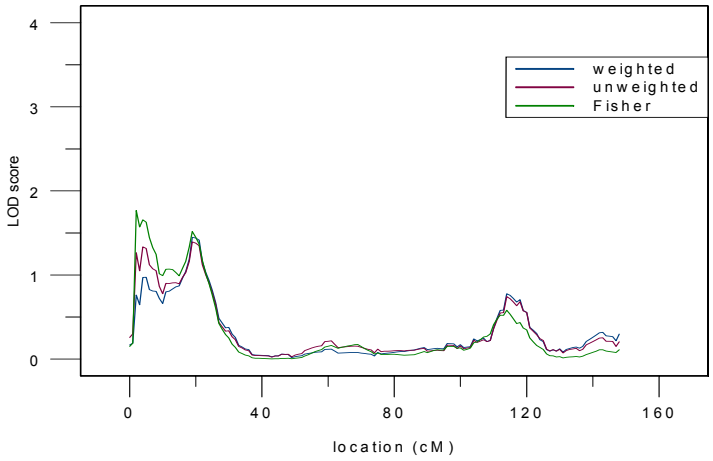
Chromosome 9



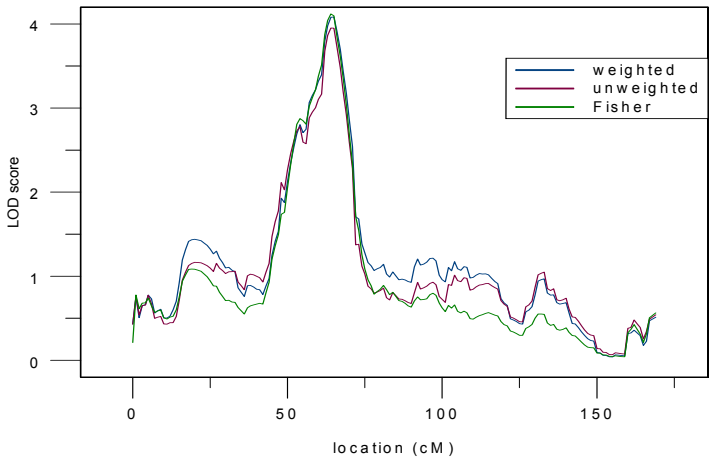
Chromosome 10



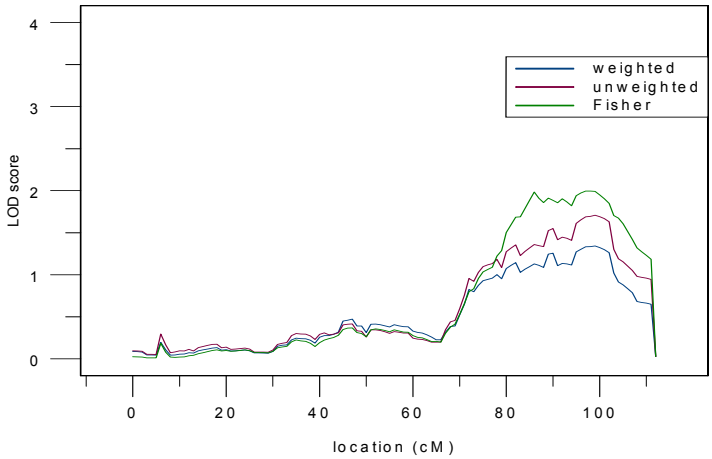
Chromosome 11



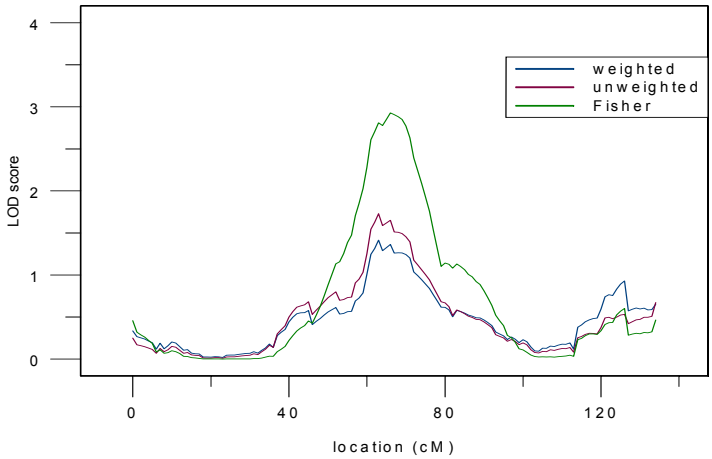
Chromosome 12



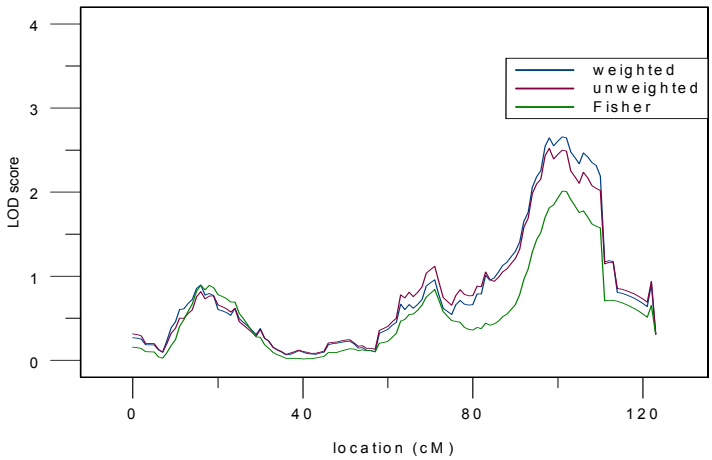
Chromosome 13



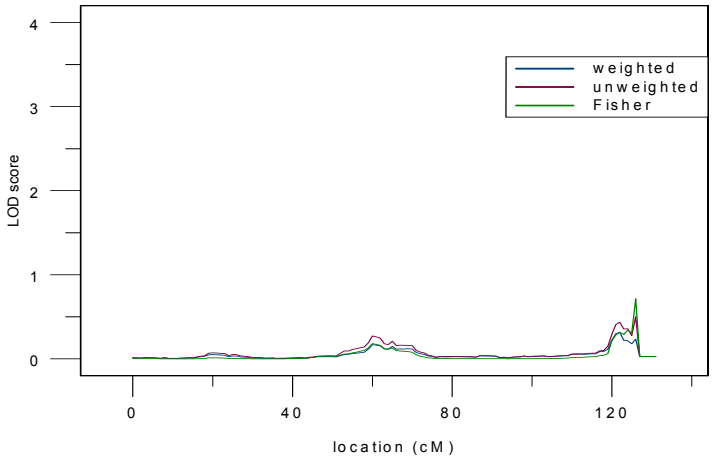
Chromosome 14



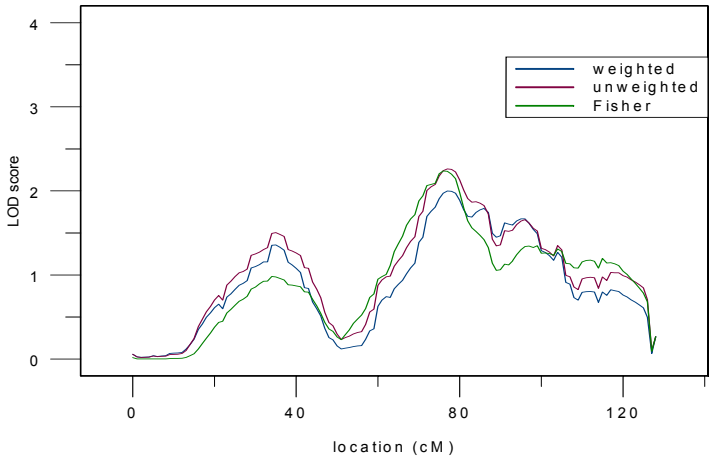
Chromosome 15



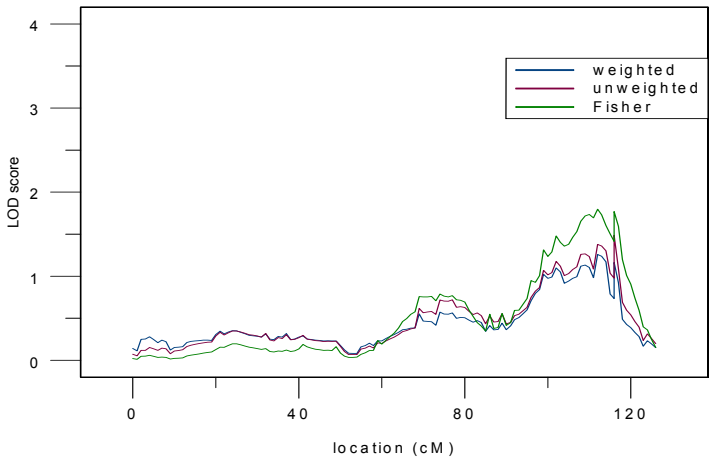
Chromosome 16



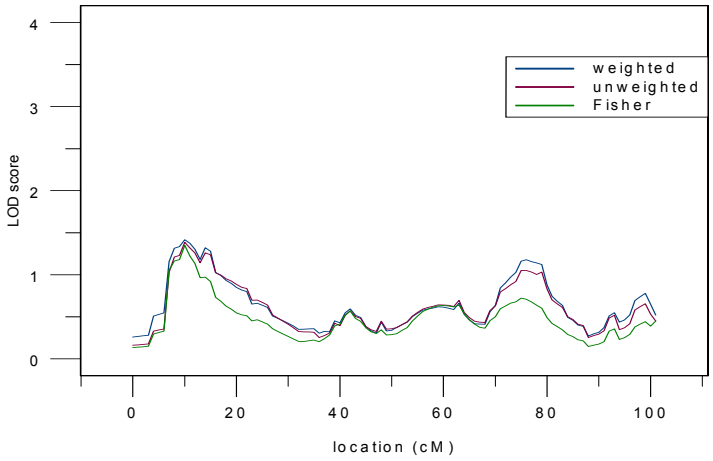
Chromosome 17



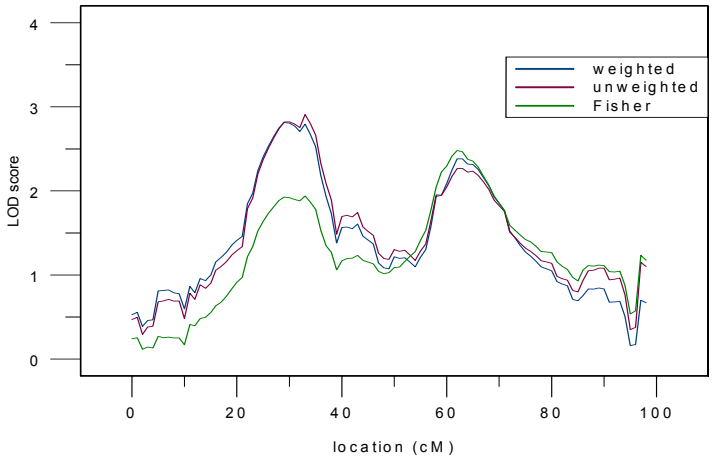
Chromosome 18



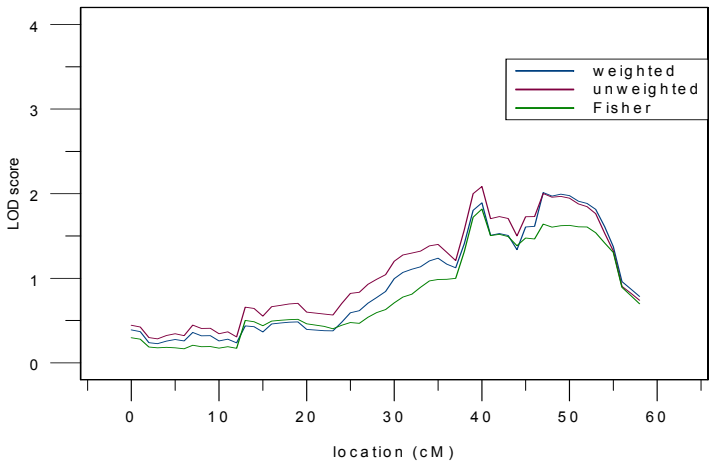
Chromosome 19

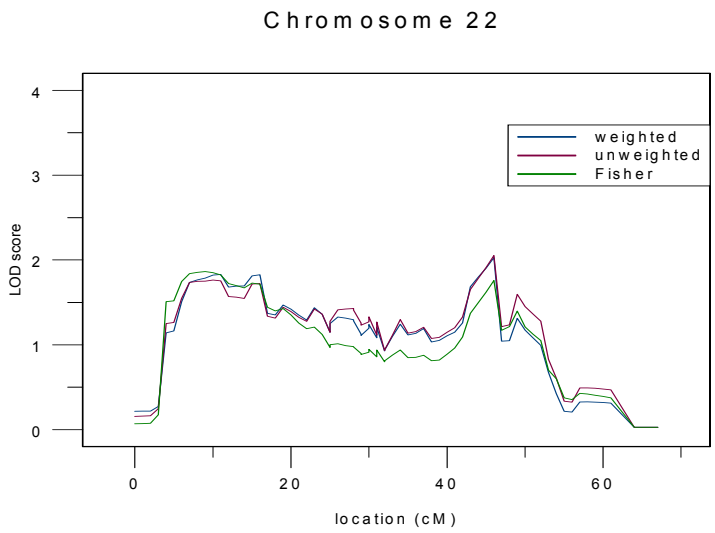


Chromosome 20



Chromosome 21





In general, for these VC LOD scores, the three methods give very similar results with the weighted and unweighted inverse normal methods even more similar and Fisher's method sometimes a bit different. The unweighted inverse normal method often gave slightly higher scores than the weighted equivalent. For the chromosomes with LOD scores above 3, Fisher's method gave the highest meta-analysis LOD scores on chromosomes 6 and 12 whereas on chromosome 7 and 9 both inverse normal methods had higher LOD scores. The highest peaks differed in magnitude by more than 1, e.g. on chromosome 6 Fisher's method gave a maximum LOD score of 4.44, while the weighted inverse normal method had only 3.33. On chromosome 7, a similar difference is seen between Fisher's method (maximum LOS score of 3.63) and the unweighted inverse normal method with 4.59.

On other chromosomes with lower meta-analysis LOD scores, even slightly larger differences exist between the methods: on chromosome 14, Fisher's method gives a maximum LOD score of 2.93 while the weighted inverse normal method reaches only 1.41.

But while the magnitude of the maxima may differ, overall the results are very similar and no method gives the highest LOD scores for the majority of regions.

The location of regional maxima is also in almost all cases identical or just 1-2 cM apart. Only on chromosome 20, the location of the maximum LOD score differs some more, because there are two local maxima (with both methods) and

the inverse normal methods reach their chromosome-wide maximum at 30 cM, while Fisher's method has the maximum at 62 cM.

6.4.2 Comparison of meta-analysis results with individual results

Figure 11 shows the LOD scores of all individual scans and the meta-analysis LOD score for Fisher's method. For ease of presentation, just one meta-analysis method is presented in these graphs.

Figure 10: Legend to figure 11.

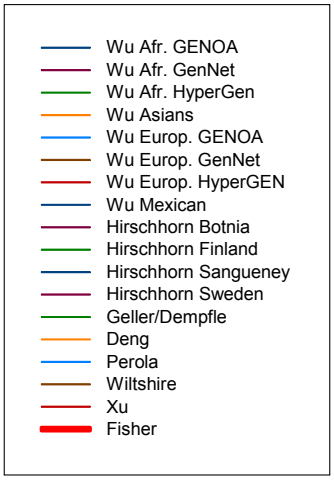
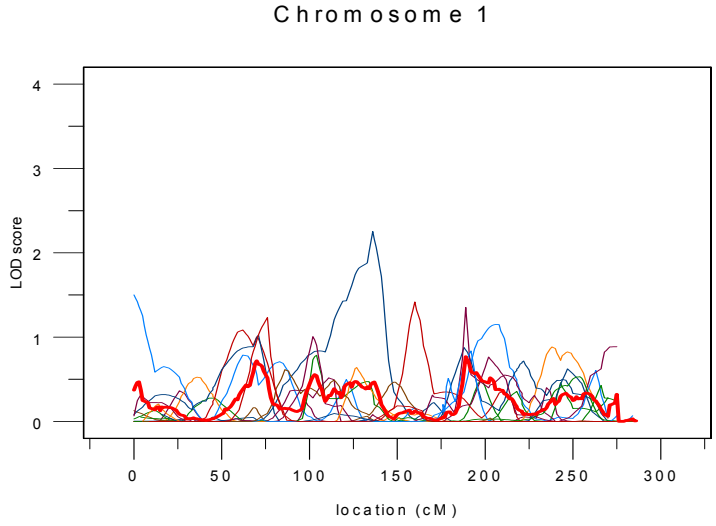
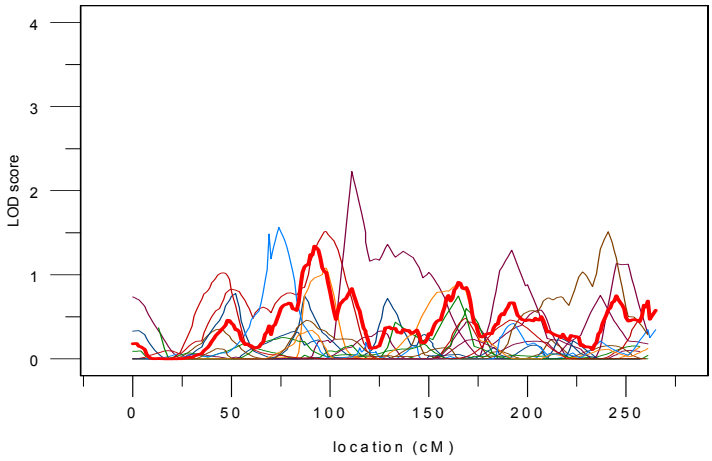


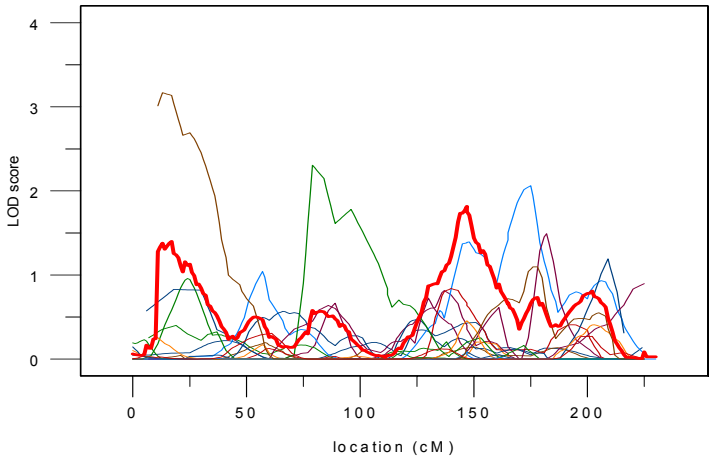
Figure 11: Comparison of LOD scores of the individual study samples (different colors, see legend above) and meta-analysis (Fisher's method, bold red line).



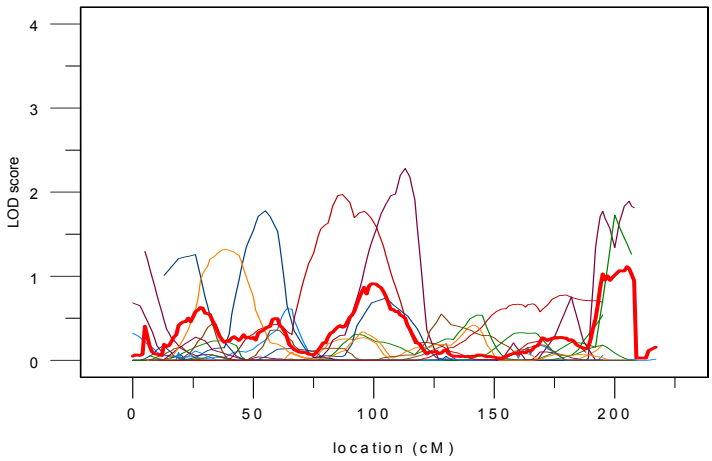
Chromosome 2



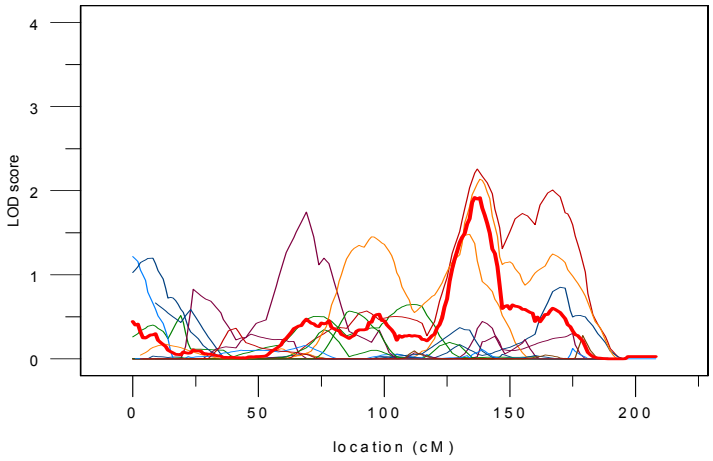
Chromosome 3



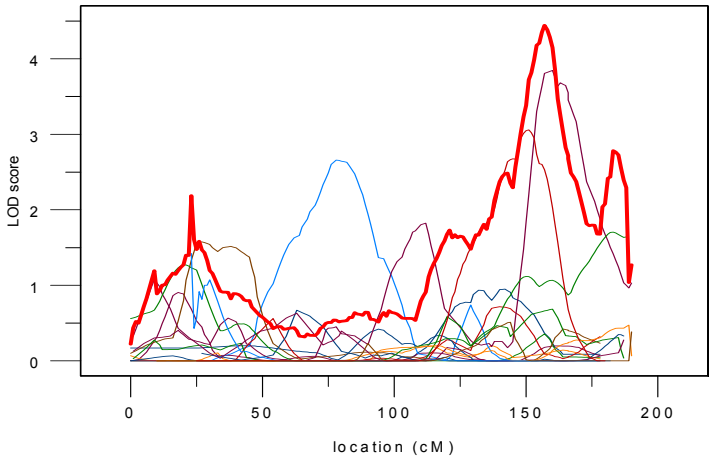
Chromosome 4



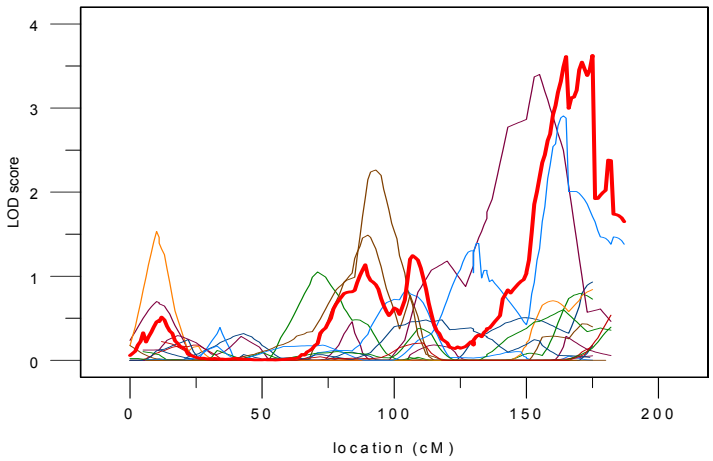
Chromosome 5



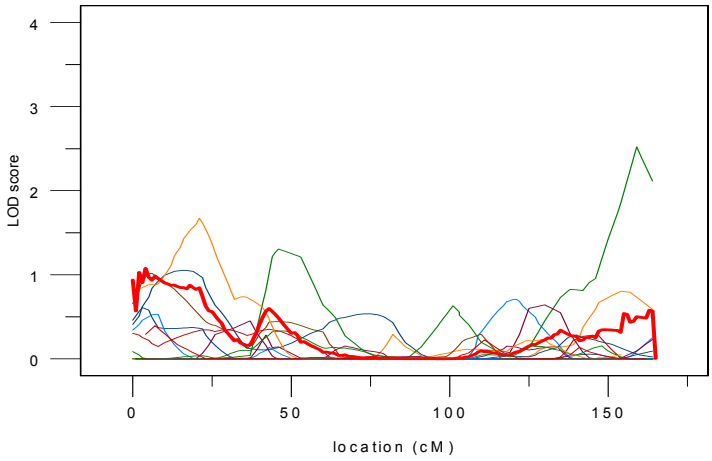
Chromosome 6



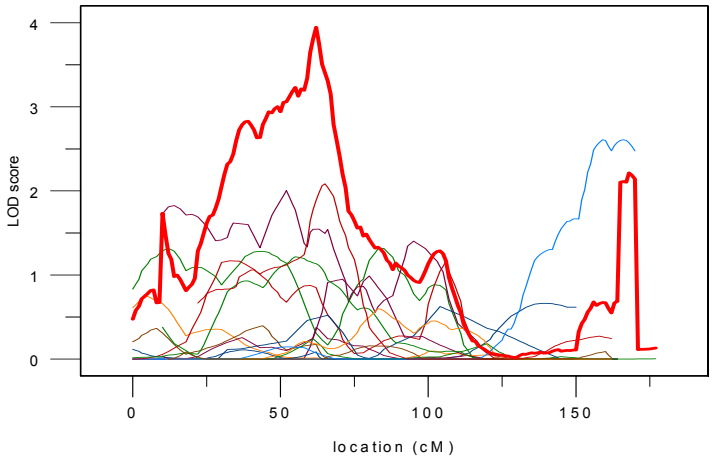
Chromosome 7



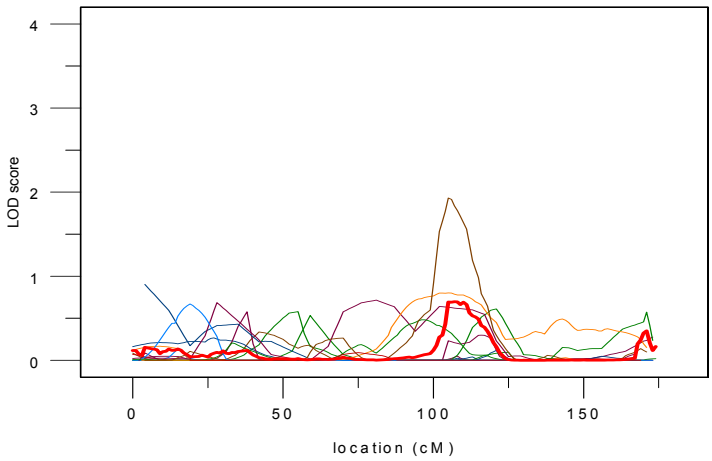
Chromosome 8



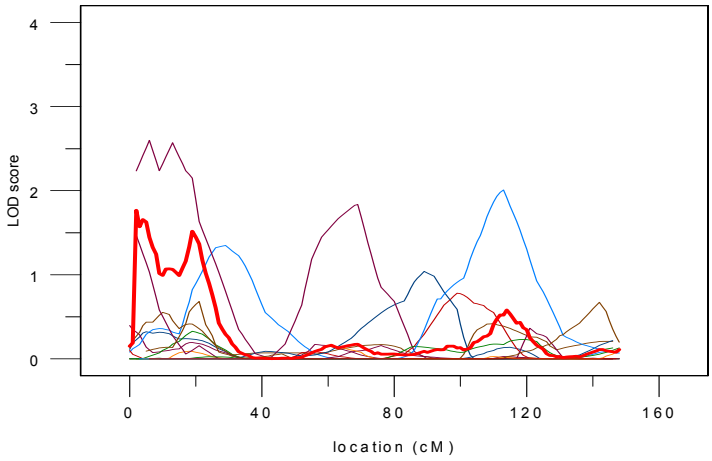
Chromosome 9



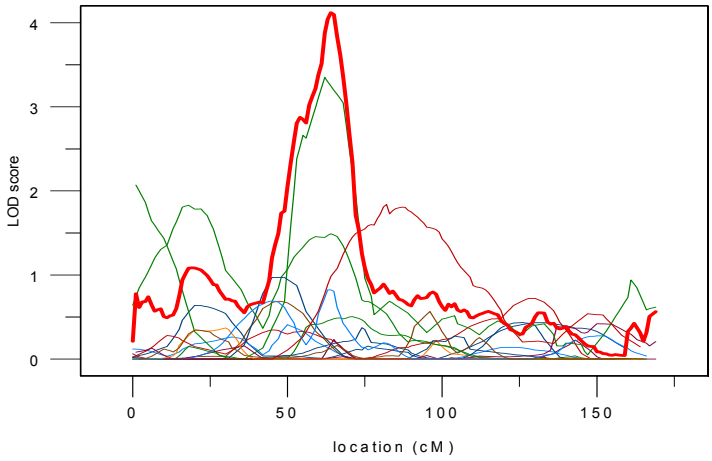
Chromosome 10



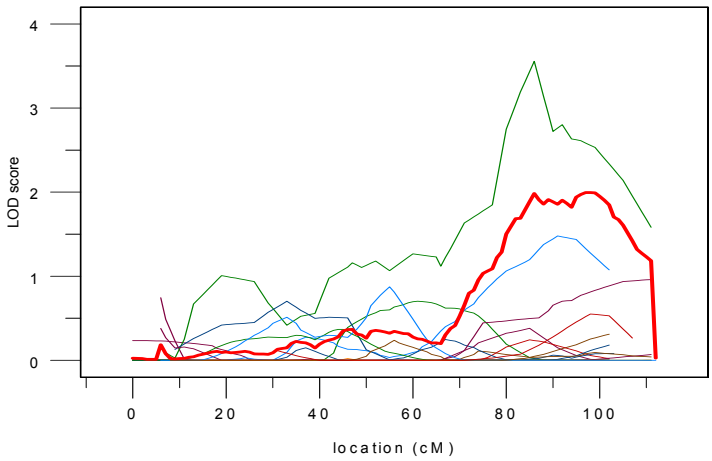
Chromosome 11



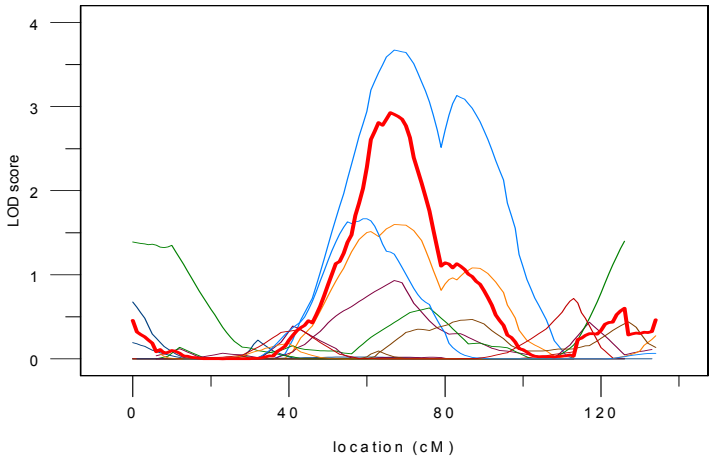
Chromosome 12



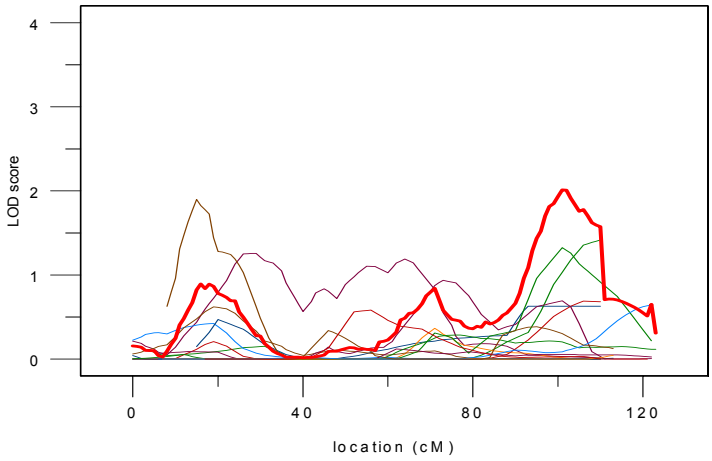
Chromosome 13



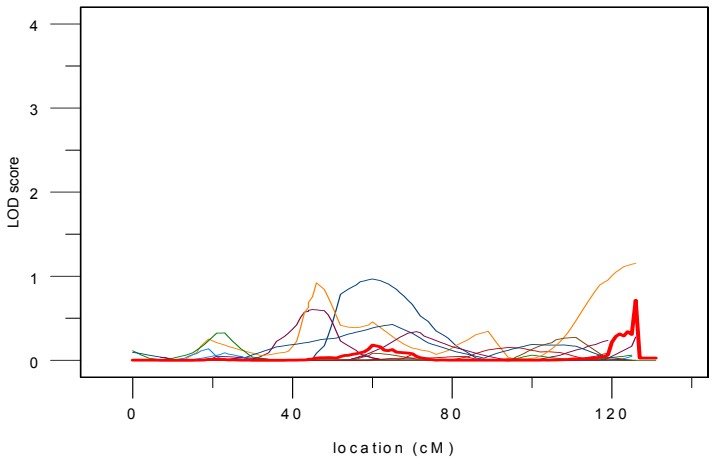
Chromosome 14



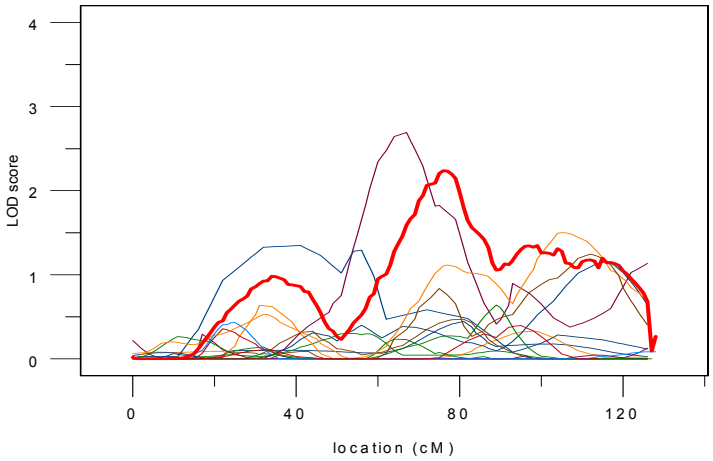
Chromosome 15



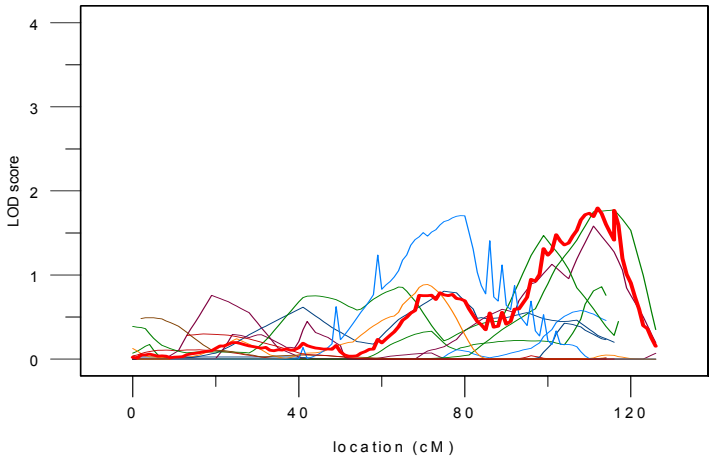
Chromosome 16



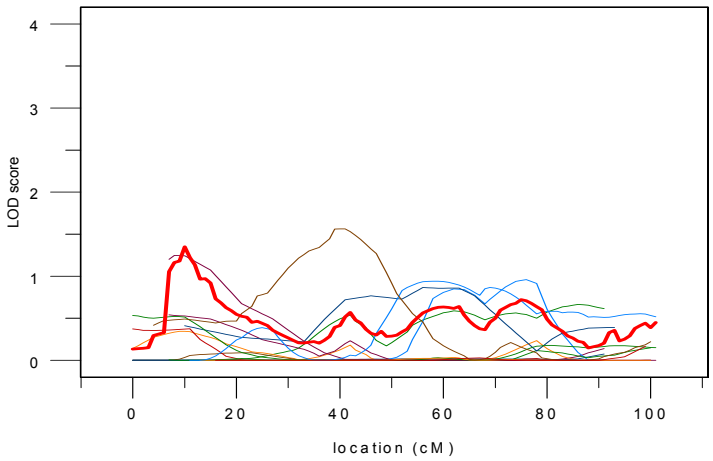
Chromosome 17



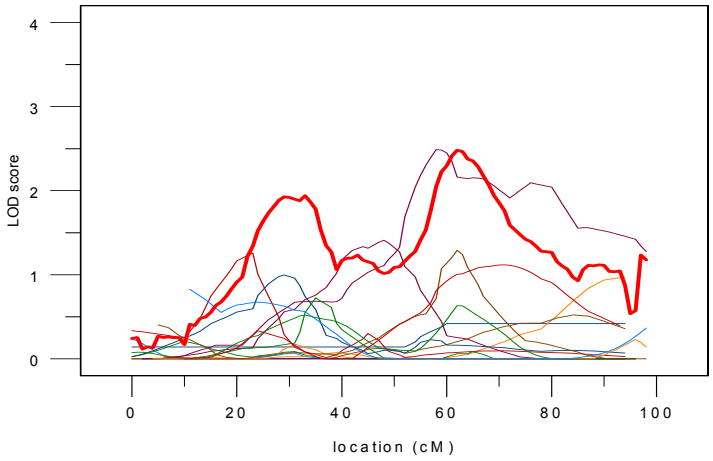
Chromosome 18



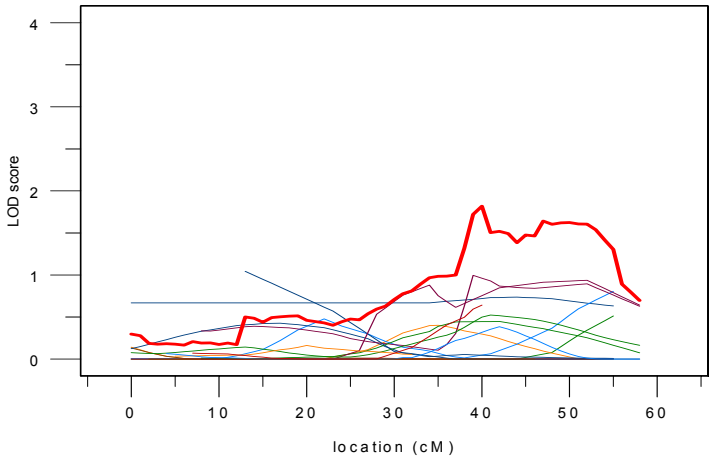
Chromosome 19



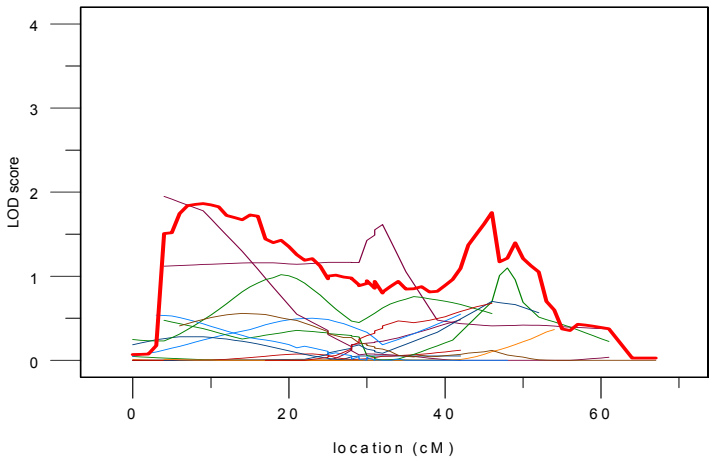
Chromosome 20



Chromosome 21



Chromosome 22



The meta-analysis yields LOD scores above 3 on chromosomes 6, 7, 9 and 12. The maximum LOD scores (Fisher's method) are 4.44 on chromosome 6 at 157 cM, 4.12 on chromosome 12 at 64 cM, 3.95 on chromosome 9 at 62 cM and 3.63 on chromosome 7 at 175 cM. The next highest LOD score is 2.93 on chromosome 14 at 66 cM.

There are several instances where interesting regions and peaks in one or more scans are confirmed by the meta-analysis. The chromosome 6 region is significant in 2 scans (Hirschhorn et al. (2001), Botnia sample and Xu et al. (2002)), but has LOD scores below 1 in the rest of the scans. Similarly for the chromosome 7 region, which was significant in the scan by Hirschhorn et al. (2001, Swedish sample) and close to significance with a LOD score of 2.91 in the scan by Perola et al. (2001), but failed to show any evidence for linkage in the other genome scans (LOD scores below 1). Even more marked is this on chromosome 12 where a significant LOD score was seen in the Finland sample (Hirschhorn et al. 2001), and only two other scans (Geller et al. 2003; Xu et al. 2002) obtained LOD scores above 1. So the regions which turn out to be significant in the meta-analysis were not considered to be "replicated" by the majority of available genome scans in the original publications.

On the other hand, some regions which show similar patterns of suggestive or even significant linkage peaks in one or more scans are not confirmed in meta-analysis. Most of the regions with suggestive linkage in only one scan are not confirmed, e.g. on chromosomes 1 and 8, and significant linkage in just one scan is often not quite substantiated by meta-analysis, e.g. on chromosomes 3 and 13, which showed evidence for linkage in the scans by Wiltshire et al. (2002) and Hirschhorn et al. (2001, Finland sample), respectively. Also regions that were suggestive in two scans such as on chromosome 5 do not turn out significant in the meta-analysis, but reached only a combined LOD score of 1.9.

And finally, the meta-analysis also reveals regions which were not significant in any scan and not identified as relevant in the individual publications. This is most notably the case on chromosome 9, where no scan gave significant results but two reached LOD scores above 2 (Hirschhorn et al. (2001, Botnia sample) and Xu et al. (2002), third and second highest peak, LOD scores 2.01 and 2.09, respectively) and several with LOD scores above 1. This region had a meta-analysis LOD score of almost 4. Xu et al. (2002) mentions the overlap with the

Botnia peak without considering the other results for this region published until then. To a lesser extent this occurred also on chromosomes 15 and 20 where all genome scans had LOD scores below 2 but the meta-analysis gives a LOD score of 2 on chromosome 15 (even 2.66 with the weighted inverse normal method) and 2.48 on chromosome 20 (2.91 with the unweighted inverse normal method). Both these regions were not among the highest in any individual scan and thus not discussed as potentially relevant. In the same region on chromosome 20, the Thompson et al. (1995) scan, which could not be included, also had its smallest p-value, so the total evidence for linkage would probably be even higher.

Another interesting aspect is the location of the meta-analysis maximum in relation to the positions of the maxima of the individual scans. In many cases, the meta-analysis reaches its maximum LOD score very close to the position of the maximum of the one or few significant individual linkage results, as can be seen nicely on chromsomes 6, 12 or 14. In some cases however, much smaller LOD scores, which are by themselves not significant, nevertheless lead to a noticable shift in the location of the meta-analysis maximum compared to the position of one prominent and dominating individual linkage peak. This is e.g. evident on chromosomes 7 or 17. Finally, in some instances, e.g. on chromosome 9, there is no dominating individual linkage peak, but rather several moderate linkage results spread over a considerable region, which leads to a broad meta-analysis peak, which all the same has a pronounced maximum.

Overall, the meta-analysis gives new insights and confirms or refutes the first impressions gained from casual inspection and comparison of individual results.

6.5 Discussion

6.5.1 Comparison of Fisher's method and weighted and unweighted inverse normal meta-analysis

The three different meta-analysis methods give very similar results in this example. Since only variance components LOD scores were available from the individual studies, these were transformed to p-values and all three applied meta-analysis methods are techniques to combine p-values of independent tests. Thus a certain similarity of results was expected. Conclusions regarding power or type 1 error cannot be drawn based on such an example with one set of real data.

6.5.2 Comparison of meta-analysis results with individual results

A formal, stringent meta-analysis has important advantages over the casual inspection of individual results or a loosely defined "replication" approach. All available data is incorporated and statistically valid results are reached. This can result in confirmation or non-confirmation of regions that showed linkage in one of the individual studies and can even reveal new linked regions, which did not display prominent linkage in any individual study.

In this meta-analysis of linkage genome scans for adult height, including 17 different samples, there are three important regions which showed linkage in one or more scans and were confirmed in the meta-analysis: the chromosome 6 region showed significant linkage in two samples, but had LOD scores below 1 in the rest of the scans. Similarly, a region on chromosome 7 previously showed one significant linkage, one study was close to significance, while the others all had LOD scores below 1 in this region. Finally on chromosome 12 again one study displayed significant linkage, and only two others reached LOD scores above 1. Using a simple replication criterion, the total of these individual results might not have been interpreted as convincing evidence for linkage. For these three regions on chromosomes 6, 7 and 12, only one or two of the 17 separate genome scans, showed strong linkage, while most scans obtained LOD scores below 1, therefore the results could easily be interpreted as false positives or as caused by substantial genetic heterogeneity between samples.

On the other hand are genetic regions which on first impression show similar patterns of individual results, but do not reach significance in the meta-analysis. This is particularly the case for most regions showing suggestive linkage in only one scan which are often not confirmed in meta-analysis, e.g. on chromosomes 1 and 8. Even regions with significant linkage in only one scan are often not confirmed in meta-analysis, e.g. on chromosome 3 (around 20 cM) or on chromosome 13.

Finally, there are some regions which did not show significant linkage in any individual scan, like the region on chromosome 9 where two genome scans showed suggestive linkage and several others had LOD scores above 1. A comparable pattern was observed on chromosome 5, where also two individual studies have suggestive linkage, but here the meta-analysis revealed an overall

not-significant result. In such cases only a formal meta-analysis allows a statistically correct interpretation of all available evidence.

6.5.3 Conclusions for the genetics of stature

The significant results of this meta-analysis of linkage genome scans of adult stature may lead the way to important insights into the genetics and molecular mechanisms of human growth. Ten chromosomal regions yielded meta-analysis LOD scores of at least 2 (and chromosome 5 had a LOD score of 1.92), four of these had LOD scores above 4 (with at least one of the meta-analysis methods). This could be compatible with the results of segregation analyses (Ginsburg et al. 1998) that found a combination of major gene and polygenic inheritance of height and goes in line with the polygenic concept that small additive effects at many loci lead to an approximately normally distributed phenotype in the general population. The results of the meta-analysis are consistent with the expectation that there are several genes with effects that are detectable in linkage analysis (albeit only in large samples). The combined sample size of genotyped and phenotyped individuals in this meta-analysis was more than 14000. This means that the power to detect genes with large effect (high locus-specific heritability) should be very high. Genetic regions with meta-analysis LOD scores around 4 might contain genes with medium locus-specific heritability, i.e. either relatively rare variants with moderate effect or rather common alleles with smaller effect, while regions with LOD scores ~2 might harbor genes with lower locus-specific heritability or no genes influencing height at all (i.e. false positives). Candidate genes for human stature in the regions identified in this meta-analysis would be expected to contribute to the normal variation in height in the general population. Additionally, there are probably many genes which are essential for growth, and as such do not show relevant variation in the normal population. In these genes, mutations might lead to extreme phenotypes (very rare monogenic diseases with severe short stature or lethal mutations) and possibly no frequent functional polymorphisms exist. This might be the case for growth hormone related genes (e.g. *GHR, GH1, GHRHR, GHSR*, see table 25) where rare mutations have been identified which lead to extremely short stature, but no common variants with subtle effect are currently known in the general population. Such genes would not be detectable with linkage studies of unselected samples, such as the population-

based studies in this meta-analysis. Accordingly, in the chromosomal regions of the *growth hormone secretagogue receptor* gene (*GHSR*, chr. 3, approximately 183 cM), the *growth hormone receptor* gene (*GHR*, chr. 5, 60 cM) and the *growth hormone releasing hormone receptor* gene (*GHRHR*, chr. 7, 49 cM) the meta-analysis LOD scores are below 0.5. Near the *growth hormone 1* gene (*GH1*, chr. 17, 82 cM) on the other hand, the meta-analysis obtained a maximum LOD score of 2.26 (at 77 cM), which might indicate that there are common polymorphisms in the *GH1* gene or its regulatory region (promoter) which influence also normal variation in stature. Different results and the identification of additional growth related loci could therefore be expected with selected samples (short or tall stature) as opposed to samples from the normal population. Similarly, the genetic analysis of variation in stature can be supplemented by the study of genetic syndromes which include short or tall stature among their cardinal features, such as Noonan Syndrome (OMIM 163950), Prader-Willi Syndrome (OMIM 176270) and many others. In some cases, genes responsible for these syndromes might also have alleles which influence normal growth variation or the whole syndrome is caused by microdeletions which include dozens of genes (e.g. Prader-Willi Syndrome), just one of which might be involved in growth regulation. E.g., the short stature seen in Léri-Weill dyschondrosteosis (LWD, OMIM 127300) and Ullrich-Turner Syndrome is caused by haploinsufficiency due to heterozygous deletions or mutations of the *SHOX* gene or its regulatory regions (Attie 2000; Benito-Sanz et al. 2005; Rao et al. 1997). Such mutations or deletions in the *SHOX* gene seem to be responsible for a fraction of patients with idiopathic short stature as well, with estimates ranging from 1% to 22% (Huber et al. 2004; Morizio et al. 2003; Rappold et al. 2002; Schneider et al. 2005; Stuppia et al. 2003) with the largest studies to date reporting only 2-2.4% (Rappold et al. 2002; Schneider et al. 2005). The *SHOX* gene is located in the pseudoautosomal region of the X- and Y-chromosome (Xp22), which was not analyzed in this meta-analysis as only Deng et al. (2002) reported X-chromosomal linkage analysis. The genetic regions of Noonan and Prader-Willi-Syndrome had meta-analysis LOD scores below 0.5, thus not supporting the possibility of relevant alleles for normal height in these regions.

Many candidate genes especially for stature and growth related phenotypes have been proposed and studied for association (see table 25 for a non-

comprehensive list of some examples). These include genes of the growth hormone-IGF-system (e.g. *GH1, GHRHR, GHSR, IGF-1*), genes regulating skeletal development and bone formation (e.g. *COL1A1, BMP2, FGFR3, VDR*), genes involved in pituitary development (e.g. *POU1F1, PROP1, LHX3, HESX1*), or genes related to sex hormones (*ESR1, CYP17, CYP19*). Studies relating to these candidate genes and monogenic forms of growth disorders have been extensively reviewed (e.g. by Palmert and Hirschhorn 2003; Castro-Feijoo et al. 2005; Kant et al. 2003).

These genes are considered candidate genes for height because of the knowledge about the gene function and their involvement in growth processes. Genome scans on the other hand have the advantage that no prior knowledge on gene function is required, so new candidates which are not obviously related to the phenotype in question can be identified. Future research should therefore investigate especially genes in linkage regions identified in this meta-analysis for their potential involvement in growth regulation. Some known candidate genes in linked regions are *BMP10* (chromosome 2), *ESR1* (chromosome 6), *VDR*, *KRAS* and *IGF-1* (all chromosome 12), *GH1* (chromosome 17) and *BMP2* (chromosome 20). These might be responsible for the observed linkage peaks in this meta-analysis. For other significant linkage regions, such as those on chromosomes 7 and 9 and smaller linkage peaks such as on chromosomes 14 and 15, no obvious candidate genes have yet been identified but careful consideration of known genes in these genetic regions might lead to plausible candidates.

Table 25: Candidate genes for height.

Chromosome	Cytogenetic region	Position (cM)	Gene name	Context	References
2	2p12	88	*Bone morphogenetic protein 10 (BMP10)*	Skeletal development	Castro-Feijoo et al. (2005)
3	3p26	19	*Basic helix-loop-helix domain containing, class B, 2 (BHLHB2, DEC1)*		Wiltshire et al. (2002)
3	3p25	37	*Peroxisome proliferative activated receptor, gamma (PPARG)*	Growth Hormone Axis	Meirhaeghe et al. (2003)
3	3p22-p21.1	69	*Parathyroid hormone receptor 1 (PTHR1)*	Skeletal development	Lei et al. (2005); Minagawa et al. (2002)
3	3q26.31	183	*Growth hormone secretagogue receptor (GHSR)*	Growth Hormone Axis	Pantel et al. (2006)
5	5p13-p12	60	*Growth hormone receptor (GHR)*	Growth Hormone Axis	Dos Santos et al. (2004)
5	5q31-q32	150	*Adrenergic, beta-2-, receptor, surface (ADRB2)*	Growth Hormone Axis	Matsuoka et al. (2002)
6	6q25.1	153	*Estrogene receptor alpha gene (ESR1)*	Sex hormones	Lorentzon et al. (2000); Schuit et al. (2004)
7	7p14	49	*Growth hormone releasing hormone receptor (GHRHR)*	Growth Hormone Axis	Castro-Feijoo et al. (2005)
7	7p12-p13	66	*Insulin-like growth factor binding protein 3 (IGFBP3)*	Growth Hormone Axis	Deal et al. (2001)
8	8p21-23	37	*Early growth response (EGR3)*		Deng et al. (2002)
10	10q24.3	125	*Cytochrome P450, family 17, subfamily A, polypeptide 1 (CYP17A1)*	Sex hormones	Zmuda et al. (2001)
11	11p15.3-p15.1	20	*Parathyroid hormone (PTH)*	Skeletal development	Lei et al. (2005)
11	11q23	106	*Dopamine D2 receptor (DRD2)*	Growth Hormone Axis	Arinami et al. (1999); Miyake et al. (1999)
12	12p12.1	45	*KRAS, Noonan Syndrom*		Schubbert et al. (2006)

Chromosome	Cytogenetic region	Position (cM)	Gene name	Context	References
12	12q12-q14	62	*Vitamin D receptor (VDR)*	Skeletal development	Suarez et al. (1997; 1998); Tao et al. (1998); Minamitani et al. (1998); Ferrara et al. (2002); van der Sluis et al. (2003)
12	12q22-q23	109	*Insulin-like growth factor 1 (IGF1)*	Growth Hormone Axis	Arends et al. (2002); Abuzzahab et al. (2003)
15	15q11-q13	8 (0-20)	*Imprinted in Prader-Willi syndrome (IPW)*		Cassidy (1997)
15	15q21.1	45	*Cytochrome P450, family 19, subfamily A, polypeptide 1 (CYP19A1)*	Sex hormones	Ellis et al. (2001)
17	17p11.2	44	*Zinc finger protein 179 (ZNF179)*		Deng et al. (2002)
17	17q21.3-q22.1	69	*Collagen type 1 alpha 1 (COL1A1)*	Skeletal development	Garnero et al. (1998); Long et al. (2004)
17	17q24.2	82	*Growth hormone 1 (GH1)*	Growth Hormone Axis	Horan et al. (2003)
17	17q24-25	101	*Growth factor receptor-bound protein 2 (GRB2)*		Deng et al. (2002)
19	19q13.32	75	*Luteinizing hormone beta (LHB)*	Sex hormones	Raivio et al. (1996)
20	20p11	18	*Bone morphogenetic protein 2 (BMP2)*	Skeletal development	Castro-Feijoo et al. (2005)
X/Y	Xp22 Pseudo-autosomal region	2	*short stature homeobox-containing gene (SHOX)*	Skeletal development	Rao et al. (1997); Rappold et al. (2002)
X	Xp22	21	*spondyloepiphyseal dysplasia, late (SEDL)*	Skeletal development	Gedeon et al. (1999)

7 General Discussion

Prospects for meta-analysis of linkage studies

For complex phenotypes, which are influenced by several genetic and environmental factors, the sample sizes achievable within one study are generally too small to yield convincing evidence for genetic linkage (Altmüller et al. 2001). However, many common complex diseases are investigated by numerous groups worldwide and meta-analysis of all available studies can result in sufficient power with combined samples (Conneally 2003). A growing number of published meta-analyses of linkage genome scans for various complex diseases has shown the increasing interest in such scientific co-operations and the good prospects for applications.

Different approaches for meta-analysis of linkage studies

Depending on the degree of detail in which data can be obtained for a meta-analysis, different statistical approaches are possible. Accordingly, the results that can be achieved by different methods of meta-analysis yield more or less detailed information about the combined effect and may include a more precise location estimate. If the complete family information, genotype and phenotype data of all samples are available, a sensible primary analysis can be performed which yields effect size estimates and their variances. These can be combined in traditional fixed or random effects meta-analysis, together with heterogeneity tests, to obtain combined effect estimates and confidence intervals which in turn give refined location estimates. If such detailed data cannot be obtained, methods such as the proposed weighted combination of NPL scores (Loesgen et al. 2001; Dempfle and Loesgen 2004), the GSMA (Wise et al. 1999) or Fisher's method (Fisher 1932) for the combination of p-values still require results for all tested markers or genetic locations, which usually will have to be requested from the authors of published genome scans. These methods do not yield combined effect estimates,

but only tests of significance, and while Fisher's method and the weighted combination of NPL scores can refine the locations of the best combined results, the GSMA gives only one p-value per 30 cM bin, thus indicating only very broad linkage regions. Only the MSP method (Badner and Gershon 2002a), a variation of Fisher's method, can be performed with just the usually published data, i.e. the smallest p-values in a genetic region with their approximate locations. Again, no effect estimation for the combined sample is possible and the resulting p-value applies to the largest region investigated in one of the original studies.

Heterogeneity

An area of further research are useful heterogeneity tests for linkage genome scans. A random effects meta-analysis, e.g. on the number of alleles shared identical by descent (IBD) between affected sib pairs (ASPs) might be advisable if heterogeneity between samples is suspected, but is only practically feasible if the detailed data such as the family specific IBD estimates are available (McQueen et al. 2005). This will usually only be the case for close collaborations where the complete raw data (family structures, genotypes, phenotypes) are accessible. Such entire data is also necessary for standard heterogeneity tests, which have low power for whole genome scans. A heterogeneity test for the Genome Search Meta-Analysis statistic (GSMA) was recently proposed using just the data commonly available for meta-analysis (Zintzaras and Ioannidis 2005), its power however has not been investigated in comparison to other approaches. If potential sources of heterogeneity, such as different ethnicities between study samples, are known, a useful strategy would be to perform meta-analyses for presumably more homogeneous subsets of samples from the same population as well as for all study samples. Additionally, methods that are explicitly designed to be powerful against specific alternatives (e.g. linkage to a certain region in only a subset of studies, i.e. only some of the null hypotheses are false) may be preferable. This includes truncation methods (variants of Fisher's method (Olkin and Saner 2001; Zaykin et al. 2002)), which consider only p-values below a certain threshold and the number of p-values above that threshold.

Proposed new meta-analysis method for linkage genome scans

A new statistical method for meta-analysis of linkage studies was proposed by Loesgen et al. (2001) which is, for non-parametric linkage analysis of ASPs, equivalent to a fixed effects meta-analysis on the number of alleles IBD. As such, study specific parameter estimates are weighted by the inverse of their variances (determined by the sample size) to obtain a combined estimate. The loss of power in non-parametric multipoint linkage analysis, which occurs for locations at which there is no fully informative marker, corresponds to a reduction in effective sample size (Guo and Elston 1999; Schork and Greenwood 2004). Therefore, weighting schemes are proposed which account for such lower information content. This becomes especially important for meta-analysis since different studies usually employ different sets of markers and information content within each study varies between genetic locations. Therefore the weighting factor for each study (representing the effective sample size) does not have to be constant throughout the genome but can vary by location.

Power comparison

The new meta-analysis method proposed by Loesgen et al. (2001) is compared to other approaches (GSMA, Fisher's method and the truncated product method) in a simulation study (Dempfle and Loesgen 2004). This shows that the incorporation of study specific characteristics (such as information content and sample size) in a meta-analysis of a qualitative trait by direct weighted combination of NPL scores has better power than previously used methods that combine p-values. Different weighting schemes for the proposed NPL score combination method are investigated and a weighting with information content and sample size has best power for the simulated scenarios. An important characteristic of the NPL combination method is that for studies with the same design (e.g. ASP studies) appropriate weighting by sample size leads to the same result as would be obtained if the raw genotype data were analyzed together. P-value methods on the other hand, are sensible to different sample sizes between studies and yield different results if the same families are split up into samples in different ways. The loss of power for non-parametric linkage methods that occurs if markers are not completely informative can be reduced by using the NPL combination method weighted by information content, if this is available for all studies.

Application of the meta-analysis method: asthma

The meta-analysis of four whole genome and five candidate region linkage studies on asthma revealed no region of significant linkage to this phenotype (Loesgen et al. 2001). The lowest p-values were obtained on chromosomes 1, 11 and 19. A possible reason for this lack of significant results is that the combined sample size is still too small, as the whole genome scans included only 340 families, and for the chromosome 5 candidate region 567 families were available. Given the moderate power of ASP linkage analysis for complex diseases, this seems to be insufficient to reliably identify genes involved in asthma susceptibility. The only asthma susceptibility gene that was yet identified by positional cloning, the ADAM33 gene (Van Eerdewegh et al. 2002), was identified in a genome scan of 460 families and it seems plausible that samples of at least this size are necessary to identify this and other asthma susceptibility genes. Different methodological issues were explored in this meta-analysis where individual data was available. Besides a demonstration of the feasibility and practical application of different weighting schemes for NPL scores, a comparison with the GSMA method was performed, and effects of marker map construction for the combined marker sets were investigated. While inaccuracies of the assumed marker map can lead to bias in multipoint linkage analysis, nowadays very precise (linkage and physical) maps are available (Kong et al. 2004; Duffy 2006) so that this does not impose serious problems for meta-analysis.

Framingham Heart Study genome scan for height

A genome scan for the quantitative phenotype adult height which was performed in a sample of extended pedigrees not ascertained for this trait, the Framingham Heart Study, yielded suggestive evidence for linkage in several genetic regions (Geller et al. 2003). Comparison of these results with previously published genome scans for the same phenotype showed remarkable concordance for some regions, as well as discordance for regions linked in other studies which were not replicated in the current sample. Thus a meta-analysis of all available genome scans for height was planned to exactly quantify the combined evidence for linkage throughout the genome.

Meta-analysis of height genome scans

The available, published genome wide linkage studies for the quantitative phenotype adult stature (Perola et al. 2001; Hirschhorn et al. 2001; Wiltshire et al. 2002; Xu et al. 2002; Deng et al. 2002; Wu et al. 2003; Geller et al. 2003) were very well suited for a meta-analysis, as almost all samples were of Caucasian origin, all samples were unselected for the phenotype of interest, the same primary statistical analysis was used for all individual studies (variance components linkage analysis) and all had used the Marshfield map as a reference map. However, only LOD scores and sample size for each study were available for the meta-analysis, so only methods appropriate for the combination of p-values could be applied and no formal heterogeneity test was possible. The three methods used (weighted and unweighted inverse normal method and Fisher's method) gave very similar results. The large combined sample size of over 14000 phenotyped and genotyped individuals enabled the identification of significant linkage to chromosomes 6, 7, 9 and 12 (with LOD scores greater 4) and revealed several "suggestive" regions (another six regions with LOD scores greater 2). This is well compatible with the results of segregation analyses which support the concept of stature as a phenotype with mostly polygenic but also major gene inheritance.

Conclusions for meta-analysis of linkage studies

The role of meta-analysis in genetic epidemiology will become more important as it is realized that for complex diseases that have a high impact on public health but are influenced by many factors with small effects, a single study alone cannot have sufficient power. While large studies are in progress, the valuable data from genome scans that were too small individually but are already completed should not be wasted but exploited efficiently. Thus, there are good prospects for future applications of meta-analysis in the area of genetic linkage studies for complex diseases, since scientists are interested in making more use of their already collected and published data to obtain higher power through increased sample sizes. Among the first efforts towards this goal, some successful meta-analyses of complex diseases have been published in the last few years (e.g. Badner and Gershon 2002b; Cox et al. 2001) and the application to the phenotype adult height presented here shows the great potentials of meta-analyses that include large combined sample sizes. The practical applicability of such meta-analyses is good

and depending on the extend of co-operation and data sharing, different methods are available. The weighted combination of NPL scores investigated here is a useful tool for research in genetic epidemiology, as it is easily applicable, and has more power than other suggested methods. It does not need raw data but leads to the same results as a pooled analysis that considers systematic differences in markers and uses only the output commonly given by model-free linkage analysis software. If less detailed results are available (such as LOD scores from variance components linkage analysis of quantitative traits), methods that are based on the combination of p-values are useful, as demonstrated in the meta-analysis of height. More research on sensible heterogeneity tests for meta-analysis of linkage genome scans would be desirable. A further increase of published meta-analyses for complex diseases can be expected in the next years.

8 Summary

Linkage genome scans for genetically complex diseases have low power with the sample sizes that were often used in the past, and hence meta-analysis of several scans for the same disease might be a promising approach. Appropriate data are now becoming accessible as many groups worldwide investigate common diseases. The aim of this thesis is to extend and evaluate statistical methodology for meta-analysis. In addition, two meta-analyses of linkage genome scans for the complex phenotypes asthma and adult stature are performed and discussed.

In the first part of this thesis, an overview of available statistical methods and current applications is given. Important differences between studies, which may lead to heterogeneity and should be accounted for in a meta-analysis, are reviewed. Some available statistical tests for heterogeneity between linkage studies are presented and their limitations for genome scans are discussed. A new meta-analysis method is introduced which is based on a weighted combination of non-parametric linkage scores. Its relationship to traditional fixed effects meta-analysis of combining parameter estimates from different studies weighted by the inverse of their respective variances is described. Recombination and low informativity of markers lead to a reduction of the effective sample size in multipoint linkage analysis. A locus specific weighting of individual studies with this effective sample size is therefore proposed. In a simulation study, the power of different methods to combine multipoint linkage scores, namely Fisher's p-value combination (Fisher 1932), the truncated product method (Zaykin et al. 2002, a variant of Fisher's method), the Genome Search Meta-Analysis (GSMA, Wise et al. 1999) method and the proposed weighting methods were compared. In particular, the effects of different genetic marker sets and sample sizes between genome scans were investigated. The weighting methods explicitly take those differences into account and have higher power in the simulated scenarios than the other methods.

The proposed meta-analysis method was applied to four linkage genome scans for the phenotype asthma and five studies of a candidate genetic region. Multipoint nonparametric linkage analysis is performed and different weighting schemes are used to combine the score statistics of individual studies to an overall statistic. For comparison, the GSMA method is also applied to the same data sets. For meta-analysis of linkage studies, a common map of genetic markers is necessary to align results obtained in different studies with different markers. In this meta-analysis, the effects of map uncertainties were evaluated. The latest versions of available combined physical and linkage maps are very precise and the small potential map errors that are left do not have relevant impact. This meta-analysis of nine asthma linkage studies does not identify significant regions of genetic linkage to asthma. A still rather small size of the combined samples may be the reason for low power to identify susceptibility genes for the complex trait asthma.

The statistical methods that can be applied for a meta-analysis of linkage studies depend crucially on the available data, especially any additional information besides the usually reported linkage statistics. For the meta-analysis of linkage genome scans for the highly heritable trait adult height, only LOD scores from variance components linkage analysis, which are measures of significance and not effect estimates, could be obtained. Thus, Fisher's method and a weighted and unweighted variant of the inverse normal method were applied. Initially, a linkage genome scan for this quantitative trait was performed in the extended pedigrees of the Framingham Heart Study. A variance components linkage analysis in this sample unselected for height gave evidence for linkage in several regions. All markers showing a LOD score greater than 1 in this analysis correspond to previously reported linkage regions, including chromosome 6q with a maximum LOD score of 2.45 and chromosomes 9, 12, 14, 18 and 22. Following this observation, a meta-analysis of all previously published genome scans for adult stature was planned. Genome scan results of 17 separate samples reported in seven publications and comprising more than 14000 phenotyped and genotyped individuals could be obtained in sufficient detail to be included in the meta-analysis. The comparison of meta-analysis results with individual studies shows that only a formal meta-analysis can exactly quantify the combined evidence for linkage and is superior to an informal classification of

results as replication or non-replication. Significant linkage of stature is observed on chromosomes 6, 7, 9 and 12 (LOD scores >4) and suggestive linkage with LOD scores >2 is obtained in six additional genetic regions. This is well compatible with the concept of height as a mostly polygenic trait for which also some major genes exist. Candidate genes in the linkage regions are discussed.

9 Zusammenfassung

Kopplungsgenomscans für genetisch komplexe Krankheiten haben mit den bislang üblichen Fallzahlen oft nur eine geringe statistische Power, daher sind Meta-Analysen von mehreren Genomscans für die gleiche Krankheit ein erfolgversprechender Ansatz. Passende Datensätze werden zunehmend verfügbar, da weltweit viele Gruppen genetische Studien zu den häufigsten Krankheiten durchführen. Ziel dieser Arbeit ist es, statistische Methoden der Meta-Analyse weiter zu entwickeln und zu evaluieren. Weiterhin werden zwei Meta-Analysen von Genomscans für komplexe Phänotypen, Asthma und Körpergröße, durchgeführt.

Im ersten Teil dieser Dissertation wird ein Überblick über aktuelle Anwendungen und bisherige statistische Methoden gegeben. Wichtige Unterschiede zwischen Studien, die zu Heterogenität führen können und in einer Meta-Analyse berücksichtigt werden sollten, werden erörtert. Einige vorhandene statistische Tests auf Heterogenität zwischen Kopplungsstudien werden vorgestellt und ihre Einschränkungen bei der Anwendung auf Genomscans diskutiert. Eine neue Methode für Meta-Analysen von genetischen Kopplungsgenomscans, die auf einer gewichteten Kombination von nicht-parametrischen Kopplungsstatistiken basiert, wird vorgestellt. Ihr Zusammenhang mit herkömmlicher „fixed-effects“ Meta-Analyse für Parameterschätzer wird erläutert. Rekombinationen und geringe Informativität genetischer Marker führen in der multipoint Kopplungsanalyse zu einer Reduzierung der effektiven Fallzahl. Eine locusspezifische Gewichtung der einzelnen Studien mit dieser effektiven Fallzahl wird vorgestellt. In einer Simulationsstudie wurde die statistische Power verschiedener Meta-Analyse Methoden für multipoint Kopplungsergebnisse verglichen. Dabei wurden die Methode nach Fisher zur Kombination von p-Werten (Fisher 1932), die „truncated product method“ (Zaykin et al. 2002, eine Variante von Fishers Methode), die Genome Search Meta-Analysis Methode

(GSMA, Wise et al. 1999) und die vorgeschlagenen Gewichtungsmethoden angewandt. Insbesondere wurden die Einflüsse unterschiedlicher genetischer Marker und Fallzahlen zwischen Genomscans untersucht. Die Gewichtungsmethoden berücksichtigen diese Unterschiede explizit und haben eine höhere statistische Power in den untersuchten Szenarien als die anderen Methoden.

Die vorgeschlagene Meta-Analyse Methode wurde auf vier Kopplungsscans und fünf Studien einer Kandidatengenregion für den Phänotyp Asthma angewandt. Zunächst wurden nicht-parametrische multipoint Kopplungsanalysen der Einzelstudien durchgeführt und die Einzel-Teststatistiken dann mit verschiedenen Gewichtungsmethoden zu einer Gesamtstatistik zusammengefasst. Zum Vergleich wurde auch die GSMA Methode auf dieselben Daten angewandt. Für eine Meta-Analyse von Kopplungsstudien benötigt man die relative genetische Position aller in den verschiedenen Studien verwendeten Marker zueinander, also eine gemeinsame genetische Karte. Die Bedeutung von Ungenauigkeiten der genetischen Karte wurde daher in dieser Studie untersucht. Die neuesten Versionen der zur Verfügung stehenden kombinierten physikalischen und genetischen Karten sind sehr präzise und die möglicherweise noch enthaltenen geringen Fehler haben keinen relevanten Einfluss auf eine Meta-Analyse. Die Meta-Analyse der neun Asthma-Studien ergab keine signifikanten Hinweise auf Kopplung. Die relativ geringe Gesamtstichprobengröße ist ein möglicher Grund für geringe statistische Power zur Identifikation von Suszeptibilitätsgenen für die genetisch komplexe Krankheit Asthma.

Welche statistischen Methoden für eine Meta-Analyse verwendet werden können, hängt stark von den zur Verfügung stehenden Daten ab, insbesondere welche weiteren Informationen neben den üblicherweise berichteten Teststatistiken vorhanden sind. Für die Meta-Analyse von Kopplungsgenomscans des Phänotyps Körpergröße standen nur LOD scores aus Varianzkomponentenanalysen zur Verfügung, welche Signifikanzmaße, nicht aber Effektstärkenschätzer sind. Daher wurden die Methode nach Fisher und eine gewichtete sowie ungewichtete Variante der Inversen-Normalverteilungsmethode angewandt. Zunächst wurde ein Kopplungsgenomscan dieses quantitativen Merkmals in den erweiterten Stammbäumen der Framingham Heart Study durchgeführt. Eine Kopplungsanalyse mit Varianzkomponentenverfahren ergab in

dieser für Körpergröße nicht speziell ausgewählten Stichprobe Kopplung zu mehreren genetischen Regionen. Alle Marker, die in dieser Auswertung einen LOD score (Kopplungsteststatistik) größer als 1 zeigen entsprechen schon früher berichteten Kopplungsregionen, darunter Chromosom 6q mit einem maximalen LOD score von 2,45 und Regionen auf den Chromsomen 9, 12, 14, 18 und 22. Auf Grund dieser Beobachtung wurde eine Meta-Analyse aller publizierten Genomscans für Körpergröße geplant. Die Ergebnisse von 17 Stichproben (aus sieben Veröffentlichungen) mit insgesamt mehr als 14000 phänotypisierten und genotypisierten Personen konnten in die Meta-Analyse einbezogen werden. Der Vergleich der Ergebnisse der Meta-Analyse mit denen der Einzelstudien zeigt, dass nur eine formale Meta-Analyse die Hinweise auf Kopplung genau quantifizieren kann und einer ungenauen Einteilung der Ergebnisse im Sinne einer Replikation oder Nicht-Replikation vorzuziehen ist. Signifikante Kopplung von Körpergröße ergibt sich zu den Chromosomen 6, 7, 9 und 12 (mit Gesamt-LOD scores >4) und Hinweise auf Kopplung mit LOD scores >2 finden sich in sechs weiteren genetischen Regionen. Dies ist gut vereinbar mit dem Konzept, dass Körpergröße hauptsächlich polygen bestimmt ist, daneben aber auch Hauptgene existieren. Schließlich werden Kandidatengene in den Kopplungsregionen diskutiert.

Parts of this doctoral thesis have been published in the following papers:

I. Loesgen S, Dempfle A, Golla A, Bickeböller H (2001). Weighting Schemes in Pooled Linkage Analysis. *Genetic Epidemiology* 21 Supplement 1: S142-S147

II. Geller F, Dempfle A, Görg T (2003). Genome Scan for BMI and height in the Framingham Heart Study. *BMC Genetics* 4 Supplement 1:S91.

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

For paper **I**, I participated in the data management and analysis, and in writing the paper, the co-authors (especially S. Loesgen) proposed the methodology and were also involved in data analysis and writing of the paper.
For paper **II**, both first authors contributed equally, i.e. they conceived the idea and design of the study, both performed the analysis and wrote the manuscript. I was responsible for the analysis of the phenotype height which is presented in this thesis, the other first author had the main responsibility for the BMI analysis, which is not included in this thesis. The third author participated in data management and prepared figures.
In case of paper **III**, both authors designed and performed the simulation study together, I had the main responsibility for writing the paper.

10 References

Abecasis G, Cox N, Daly MJ, Kruglyak L, Laird N, Markianos K, and Patterson N (2004) No bias in linkage analysis. *Am J Hum Genet* 75:722-723

Abecasis GR, Cherny SS, Cookson WO, and Cardon LR (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97-101

Abuzzahab MJ, Schneider A, Goddard A et al (2003) IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. *N Engl J Med* 349:2211-2222

Allison DB and Heo M (1998) Meta-analysis of linkage data under worst-case conditions: a demonstration using the human OB region. *Genetics* 148:859-865

Almasy L and Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62:1198-1211

Altmüller J, Palmer LJ, Fischer G, Scherb H, and Wjst M (2001) Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 69:936-950

An P, Freedman BI, Hanis CL, Chen YD, Weder AB, Schork NJ, Boerwinkle E, Province MA, Hsiung CA, Wu X, Quertermous T, and Rao DC (2005) Genome-wide linkage scans for fasting glucose, insulin, and insulin resistance in the National Heart, Lung, and Blood Institute Family Blood Pressure Program: evidence of linkages to chromosome 7q36 and 19q13 from meta-analysis. *Diabetes* 54:909-914

Arends N, Johnston L, Hokken-Koelega A, van Duijn C, de Ridder M, Savage M, and Clark A (2002) Polymorphism in the IGF-I gene: clinical relevance for short children born small for gestational age (SGA). *J Clin Endocrinol Metab* 87:2720

Arinami T, Iijima Y, Yamakawa-Kobayashi K, Ishiguro H, Ohtsuki T, Yanagi H, Shimakura Y, Ishikawa H, and Hamaguchi H (1999) Supportive evidence for contribution of the dopamine D2 receptor gene to heritability of stature: linkage and association studies. *Ann Hum Genet* 63:147-151

Attie KM (2000) Genetic studies in idiopathic short stature. *Curr Opin Pediatr* 12:400-404

Babron MC, Selinger-Leneman H, Dizier MH, and Clerget-Darpoux F (2001) Homogeneity of asthma genome scan results. *Genet Epidemiol* 21 Suppl 1:S44-S48

Badner JA and Gershon ES (2002a) Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* 7:405-411

Badner JA and Gershon ES (2002b) Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7. *Mol Psychiatry* 7:56-66

Badner JA and Goldin LR (1999) Meta-analysis of linkage studies. *Genet Epidemiol* 17 Suppl 1:S485-S490

Benito-Sanz S, Thomas NS, Huber C, Gorbenco del Blanco D, Aza-Carmona M, Crolla JA, Maloney V, Rappold G, Argente J, Campos-Barros A, Cormier-Daire V, and Heath KE (2005) A Novel Class of Pseudoautosomal Region 1 Deletions Downstream of SHOX Is Associated with Leri-Weill Dyschondrosteosis. *Am J Hum Genet* 77:533-544

Blackwelder WC and Elston RC (1985) A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet Epidemiol* 2:85-97

Blakey J, Halapi E, Bjornsdottir US, Wheatley A, Kristinsson S, Upmanyu R, Stefansson K, Hakonarson H, and Hall IP (2005) Contribution of ADAM33 polymorphisms to the population risk of asthma. *Thorax* 60:274-276

Botstein D, White RL, Skolnick M, and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331

Bravata DM and Olkin I (2001) Simple pooling versus combining in meta-analysis. *Eval Health Prof* 24:218-230

Brazma A, Hingamp P, Quackenbush J et al (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 29:365-371

Broman KW, Murray JC, Sheffield VC, White RL, and Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* 63:861-869

Cassidy SB (1997) Prader-Willi syndrome. *J Med Genet* 34:917-923

Castro-Feijoo L, Quinteiro C, Loidi L, Barreiro J, Cabanas P, Arevalo T, Dieguez C, Casanueva FF, and Pombo M (2005) Genetic basis of short stature. *J Endocrinol Invest* 28:30-37

Chen WM and Deng HW (2001) A general and accurate approach for computing the statistical power of the transmission disequilibrium test for complex disease genes. *Genet Epidemiol* 21:53-67

Choi SJ, Rho YH, Ji JD, Song GG, and Lee YH (2006) Genome scan meta-analysis of rheumatoid arthritis. *Rheumatology (Oxford)* 45:166-170

Cochran WG (1937) Problems arising in the analysis of a series of similar experiments. *J Roy Statist Soc* 4:102-118

Conneally PM (2003) The complexity of complex diseases. *Am J Hum Genet* 72:229-232

Cordell HJ (2001) Sample size requirements to control for stochastic variation in magnitude and location of allele-sharing linkage statistics in affected sibling pairs. *Ann Hum Genet* 65:491-502

Cordell HJ (2004) Bias toward the null hypothesis in model-free linkage analysis is highly dependent on the test statistic used. *Am J Hum Genet* 74:1294-1302

Cox NJ, Wapelhorst B, Morrison VA, Johnson L, Pinchuk L, Spielman RS, Todd JA, and Concannon P (2001) Seven regions of the genome show evidence of linkage to type 1 diabetes in a consensus analysis of 767 multiplex families. *Am J Hum Genet* 69:820-830

Daniels SE, Bhattacharrya S, James A, Leaves NI, Young A, Hill MR, Faux JA, Ryan GF, le Souef PN, Lathrop GM, Musk AW, and Cookson WO (1996) A genome-wide search for quantitative trait loci underlying asthma. *Nature* 383:247-250

Daw EW, Thompson EA, and Wijsman EM (2000) Bias in multipoint linkage analysis arising from map misspecification. *Genet Epidemiol* 19:366-380

Deal C, Ma J, Wilkin F, Paquette J, Rozen F, Ge B, Hudson T, Stampfer M, and Pollak M (2001) Novel promoter polymorphism in insulin-like growth factor-binding protein-3: correlation with serum levels and interaction with known regulators. *J Clin Endocrinol Metab* 86:1274-1280

Demenais F, Kanninen T, Lindgren CM, Wiltshire S, Gaget S, Dandrieux C, Almgren P, Sjogren M, Hattersley A, Dina C, Tuomi T, McCarthy MI, Froguel P, and Groop LC (2003) 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. *Hum Mol Genet* 12:1865-1873

Dempfle A and Loesgen S (2004) Meta-analysis of linkage studies for complex diseases: an overview of methods and a simulation study. *Ann Hum Genet* 68:69-83

Deng HW, Xu FH, Liu YZ, Shen H, Deng H, Huang QY, Liu YJ, Conway T, Li JL, Davies KM, and Recker RR (2002) A whole-genome linkage scan suggests several genomic regions potentially containing QTLs underlying the variation of stature. *Am J Med Genet* 113:29-39

Dos Santos C, Essioux L, Teinturier C, Tauber M, Goffin V, and Bougneres P (2004) A common polymorphism of the growth hormone receptor is associated with increased responsiveness to growth hormone. *Nat Genet* 36:720-724

Duffy DL (2006) An integrated genetic map for linkage analysis. *Behav Genet* 36:4-6

Ellis JA, Stebbing M, and Harrap SB (2001) Significant population variation in adult male height associated with the Y chromosome and the aromatase gene. *J Clin Endocrinol Metab* 86:4147-4150

Elston RC (1991) On Fisher's Method of Combining *p*-Values. *Biom J* 33:339-345

Etzel CJ and Costello TJ (2001) Assessing linkage of immunoglobulin E using a meta-analysis approach. *Genet Epidemiol* 21 Suppl 1:S97-102

Etzel CJ and Guerra R (2002) Meta-analysis of Genetic-Linkage Analysis of Quantitative-Trait Loci. *Am J Hum Genet* 71:56-65

Ferrara M, Matarese SM, Francese M, Borrelli B, Coppola A, Coppola L, and Esposito L (2002) Effect of VDR polymorphisms on growth and bone mineral density in homozygous beta thalassaemia. *Br J Haematol* 117:436-440

Fisher RA. 1932. Statistical Methods for Research Workers. 4 ed. Edinburgh: Oliver and Boyd

Fisher SA, Abecasis GR, Yashar BM et al (2005) Meta-analysis of genome scans of age-related macular degeneration. *Hum Mol Genet* 14:2257-2264

Folks LJ (1984) Combination of Independent Tests. In: Krishnaiah PR, Sen PK (eds) Handbook of Statistics, vol. 4. Amsterdam: Elsevier Science

Franke D and Ziegler A (2005) Weighting affected sib pairs by marker informativity. *Am J Hum Genet* 77:230-241 (Erratum in: *Am J Hum Genet*. (2006) 78:732.)

Fulker DW, Cherny SS, and Cardon LR (1995) Multipoint interval mapping of quantitative trait loci, using sib pairs. *Am J Hum Genet* 56:1224-1233

Garnero P, Borel O, Grant SF, Ralston SH, and Delmas PD (1998) Collagen Ialpha1 Sp1 polymorphism, bone mass, and bone turnover in healthy French premenopausal women: the OFELY study. *J Bone Miner Res* 13:813-817

Gedeon AK, Colley A, Jamieson R, Thompson EM, Rogers J, Sillence D, Tiller GE, Mulley JC, and Gecz J (1999) Identification of the gene (SEDL) causing X-linked spondyloepiphyseal dysplasia tarda. *Nat Genet* 22:400-404

Geller F, Dempfle A, and Görg T (2003) Genome Scan for BMI and height in the Framingham Heart Study. *BMC Genetics* 4 Suppl 1:S91

Ginsburg E, Livshits G, Yakovenko K, and Kobyliansky E (1998) Major gene control of human body height, weight and BMI in five ethnically different populations. *Ann Hum Genet* 62:307-322

Goldstein DR, Sain SR, Guerra R, and Etzel CJ (1999) Meta-analysis by combining parameter estimates: simulated linkage studies. *Genet Epidemiol* 17 Suppl 1:S581-S586

Good IJ (1955) On the weighted combination of significance tests. *Journal of the Royal Statistical Society, Series B* 17:264-265

Göring HH, Terwilliger JD, and Blangero J (2001) Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet* 69:1357-1369

Greenberg DA, MacCluer JW, Spence MA, Falk CT, and Hodge SE (1999) Simulated data for a complex genetic trait (problem 2 for GAW11): how the model was developed, and why. *Genet Epidemiol* 17 Suppl 1:S449-S459

Gu C, Province M, and Rao DC (1999) Meta-analysis of genetic linkage to quantitative trait loci with study- specific covariates: a mixed-effects model. *Genet Epidemiol* 17 Suppl 1:S599-S604

Gu C, Province M, Todorov A, and Rao DC (1998) Meta-analysis methodology for combining non-parametric sibpair linkage results: genetic homogeneity and identical markers. *Genet Epidemiol* 15:609-626

Gu C, Province MA, Rao DC (2001) Meta-analysis for model-free methods. In: Rao DC, Province MA (eds) Genetic Dissection of Complex Traits*,* 42 ed. San Diego: Academic Press

Gudbjartsson DF, Jonasson K, Frigge ML, and Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 25:12-13

Guerra R, Etzel CJ, Goldstein DR, and Sain SR (1999) Meta-analysis by combining p-values: simulated linkage studies. *Genet Epidemiol* 17 Suppl 1:S605-S609

Guo X and Elston RC (1999) Linkage information content of polymorphic genetic markers. *Hum Hered* 49:112-118

Guo X and Elston RC (2000) Two-stage global search designs for linkage analysis I: use of the mean statistic for affected sib pairs. *Genet Epidemiol* 18:97-110

Guo X, Olson JM, Elston RC, and Niu T (2002) The linkage information content value of polymorphism genetic markers in model-free linkage analysis. *Hum Hered* 53:45-48

Halpern J and Whittemore AS (1999) Multipoint linkage analysis. A cautionary note. *Hum Hered* 49:194-196

Haseman JK and Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2:3-19

Hedges LV, Olkin I. 1985. Statistical Methods for Meta-Analysis*.* San Diego: Academic Press

Heijmans BT, Beekman M, Putter H, Lakenberg N, van der Wijk HJ, Whitfield JB, Posthuma D, Pedersen NL, Martin NG, Boomsma DI, and Slagboom PE (2005) Meta-analysis of four new genome scans for lipid parameters and analysis

of positional candidates in positive linkage regions. *Eur J Hum Genet* 13:1143-1153

Hirschhorn JN, Lindgren CM, Daly MJ, Kirby A, Schaffner SF, Burtt NP, Altshuler D, Parker A, Rioux JD, Platko J, Gaudet D, Hudson TJ, Groop LC, and Lander ES (2001) Genomewide linkage analysis of stature in multiple populations reveals several regions with evidence of linkage to adult height. *Am J Hum Genet* 69:106-116

Holgate ST, Yang Y, Haitchi HM, Powell RM, Holloway JW, Yoshisue H, Pang YY, Cakebread J, and Davies DE (2006) The genetics of asthma: ADAM33 as an example of a susceptibility gene. *Proc Am Thorac Soc* 3:440-443

Holmans P (1993) Asymptotic properties of affected-sib-pair linkage analysis. *Am J Hum Genet* 52:362-374

Horan M, Millar DS, Hedderich J, Lewis G, Newsway V, Mo N, Fryklund L, Procter AM, Krawczak M, and Cooper DN (2003) Human growth hormone 1 (GH1) gene expression: complex haplotype-dependent influence of polymorphic variation in the proximal promoter and locus control region. *Hum Mutat* 21:408-423

Hsu LM (2005) Some properties of r equivalent: a simple effect size indicator. *Psychol Methods* 10:420-427

Huber C, Ipsas-Jouron S, Rossilio M, Salaun-Martin C, Munnich A, and Cormier-Daire V (2004) Molecular analysis of the SHOX gene in a series of 100 patients with short stature. Paper presented at the *54th Anual Meeting of the American Society of Human Genetics* abstract 187

Iyengar SK, Jacobs KB, and Palmer LJ (2001) Improved evidence for linkage on 6p and 5p with retrospective pooling of data from three asthma genome screens. *Genet Epidemiol* 21 Suppl 1:S130-S135

Jacobs KB, Burton PR, Iyengar SK, Elston RC, and Palmer LJ (2001) Pooling data and linkage analysis in the chromosome 5q candidate region for asthma. *Genet Epidemiol* 21 Suppl 1:S103-S108

Johnson L, Luke A, Adeyemo A, Deng HW, Mitchell BD, Comuzzie AG, Cole SA, Blangero J, Perola M, and Teare MD (2005) Meta-analysis of five genome-wide linkage studies for body mass index reveals significant evidence for linkage to chromosome 8p. *Int J Obes* 29:413-419

Kant SG, Wit JM, and Breuning MH (2003) Genetic analysis of short stature. *Horm Res* 60:157-165

Knapp M (2006) Power comparison of generalizations of the mean test for affected sib pairs in case of incompletely informative markers. *Genet Epidemiol* 30:314-319

Koivukoski L, Fisher SA, Kanninen T, Lewis CM, von Wowern F, Hunt S, Kardia SL, Levy D, Perola M, Rankinen T, Rao DC, Rice T, Thiel BA, and Melander O (2004) Meta-analysis of genome-wide scans for hypertension and blood pressure in Caucasians shows evidence of susceptibility regions on chromosomes 2 and 3. *Hum Mol Genet* 13:2325-2332

Kong A and Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 61:1179-1188

Kong A, Gudbjartsson DF, Sainz J et al (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31:241-247

Kong X, Murphy K, Raj T, He C, White PS, and Matise TC (2004) A combined linkage-physical map of the human genome. *Am J Hum Genet* 75:1143-1148

Kraemer HC (2005) A simple effect size indicator for two-group comparisons? A comment on r equivalent. *Psychol Methods* 10:413-419

Krawczak M (2001) ASP-a simulation-based power calculator for genetic linkage studies of qualitative traits, using sib-pairs. *Hum Genet* 109:675-677

Kruglyak L, Daly MJ, and Lander ES (1995) Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. *Am J Hum Genet* 56:519-527

Kruglyak L, Daly MJ, Reeve-Daly MP, and Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347-1363

Kruglyak L and Lander ES (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57:439-454

Lander ES and Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241-247

Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921

Lander ES and Schork NJ (1994) Genetic dissection of complex traits. *Science* 265:2037-2048

Lathrop GM, Lalouel JM, Julier C, and Ott J (1984) Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci U S A* 81:3443-3446

Lei SF, Wang YB, Liu MY, Mo XY, and Deng HW (2005) The VDR, COL1A1, PTH, and PTHR1 gene polymorphisms are not associated with bone size and height in Chinese nuclear families. *J Bone Miner Metab* 23:501-505

Levinson DF, Holmans PA, Laurent C et al (2002) No major schizophrenia locus detected on chromosome 1q in a large multicenter sample. *Science* 296:739-741

Li Z and Rao DC (1996) Random effects model for meta-analysis of multiple quantitative sibpair linkage studies. *Genet Epidemiol* 13:377-383

Liu YZ, Xu FH, Shen H et al (2004) Genetic dissection of human stature in a large sample of multiplex pedigrees. *Ann Hum Genet* 68:472-488

Loesgen S, Dempfle A, Golla A, and Bickeböller H (2001) Weighting schemes in pooled linkage analysis. *Genet Epidemiol* 21 Suppl 1:S142-S147

Long JR, Liu PY, Lu Y, Xiong DH, Zhao LJ, Zhang YY, Elze L, Recker RR, and Deng HW (2004) Association between COL1A1 gene polymorphisms and bone size in Caucasians. *Eur J Hum Genet* 12:383-388

Lonjou C, Barnes K, Chen H, Cookson WO, Deichmann KA, Hall IP, Holloway JW, Laitinen T, Palmer LJ, Wjst M, and Morton NE (2000) A first trial of retrospective collaboration for positional cloning in complex inheritance: assay of the cytokine region on chromosome 5 by the consortium on asthma genetics (COAG). *Proc Natl Acad Sci U S A* 97:10942-10947

Lorentzon M, Lorentzon R, and Nordstrom P (2000) Vitamin D receptor gene polymorphism is associated with birth height, growth to adolescence, and adult stature in healthy caucasian men: a cross-sectional and longitudinal study. *J Clin Endocrinol Metab* 85:1666-1670

Malhotra A, Coon H, Feitosa MF, Li WD, North KE, Price RA, Bouchard C, Hunt SC, and Wolford JK (2005) Meta-analysis of genome-wide linkage studies for quantitative lipid traits in African Americans. *Hum Mol Genet* 14:3955-3962

Marazita ML, Murray JC, Lidral AC et al (2004) Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32-35. *Am J Hum Genet* 75:161-173

Matsuoka H, Iwama S, Miura N, Ikezaki A, and Sugihara S (2002) Impact of polymorphisms of human beta-adrenergic receptor gene on changes in height during growth hormone treatment. *Endocr J* 49:21-28

McQueen MB, Blacker D, and Laird N (2006) Variance Calculations for Identity-by-Descent Estimation. *Am J Hum Genet* 78:914-921

McQueen MB, Devlin B, Faraone SV et al (2005) Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *Am J Hum Genet* 77:582-595

Meirhaeghe A, Fajas L, Gouilleux F, Cottel D, Helbecque N, Auwerx J, and Amouyel P (2003) A functional polymorphism in a STAT5B site of the human PPAR gamma 3 gene promoter affects height and lipid metabolism in a French population. *Arterioscler Thromb Vasc Biol* 23:289-294

Meyers DA, Wjst M, and Ober C (2001) Description of three data sets: Collaborative Study on the Genetics of Asthma (CSGA), the German Affected-

Sib-Pair Study, and the Hutterites of South Dakota. *Genet Epidemiol* 21 Suppl 1:S4-S8

Minagawa M, Yasuda T, Watanabe T, Minamitani K, Takahashi Y, Goltzman D, White JH, Hendy GN, and Kohno Y (2002) Association between AAAG repeat polymorphism in the P3 promoter of the human parathyroid hormone (PTH)/PTH-related peptide receptor gene and adult height, urinary pyridinoline excretion, and promoter activity. *J Clin Endocrinol Metab* 87:1791-1796

Minamitani K, Takahashi Y, Minagawa M, Yasuda T, and Niimi H (1998) Difference in height associated with a translation start site polymorphism in the vitamin D receptor gene. *Pediatr Res* 44:628-632

Miyake H, Nagashima K, Onigata K, Nagashima T, Takano Y, and Morikawa A (1999) Allelic variations of the D2 dopamine receptor gene in children with idiopathic short stature. *J Hum Genet* 44:26-29

Morizio E, Stuppia L, Gatta V, Fantasia D, Guanciali FP, Rinaldi MM, Scarano G, Concolino D, Giannotti A, Verrotti A, Chiarelli F, Calabrese G, and Palka G (2003) Deletion of the SHOX gene in patients with short stature of unknown cause. *Am J Med Genet A* 119:293-296

Morton NE (1955) Sequential Test for the Detection of Linkage. *Am J Hum Genet* 7:277-318

Mukhopadhyay I, Feingold E, Wang T, Elston RC, and Weeks DE (2006) Treatment of Uninformative Families in Mean Allele Sharing Tests for Linkage. *Statistical Applications in Genetics and Molecular Biology* 5:Article 13. Available at: http://www.bepress.com/sagmb/vol5/iss1/art13

Mukhopadhyay I, Feingold E, and Weeks DE (2004) No "bias" toward the null hypothesis in most conventional multipoint nonparametric linkage analyses. *Am J Hum Genet* 75:716-718

Nair RP, Henseler T, Jenisch S, Stuart P, Bichakjian CK, Lenk W, Westphal E, Guo SW, Christophers E, Voorhees JJ, and Elder JT (1997) Evidence for two

psoriasis susceptibility loci (HLA and 17q) and two novel candidate regions (16q and 20p) by genome-wide scan. *Hum Mol Genet* 6:1349-1356

Nyholt DR (2000) All LODs are not created equal. *Am J Hum Genet* 67:282-288

Ober C, Tsalenko A, Parry R, and Cox NJ (2000) A second-generation genomewide screen for asthma-susceptibility alleles in a founder population. *Am J Hum Genet* 67:1154-1162

Olkin I and Saner H (2001) Approximations for trimmed Fisher procedures in research synthesis. *Stat Methods Med Res* 10:267-276

Palmer LJ, Cookson WO, Deichmann KA, Holloway JW, and Laitinen T (2001) Single region linkage analyses of asthma: description of data sets. *Genet Epidemiol* 21 Suppl 1:S9-15

Palmert MR and Hirschhorn JN (2003) Genetic approaches to stature, pubertal timing, and other complex traits. *Mol Genet Metab* 80:1-10

Pantel J, Legendre M, Cabrol S, Hilal L, Hajaji Y, Morisset S, Nivot S, Vie-Luton MP, Grouselle D, de Kerdanet M, Kadiri A, Epelbaum J, Le Bouc Y, and Amselem S (2006) Loss of constitutive activity of the growth hormone secretagogue receptor in familial short stature. *J Clin Invest* 116:760-768

Pearson K and Lee A (1903) On the laws of inheritance in man. I. Inheritance of physical characters. *Biometrika* 2:357-462

Perola M, Ohman M, Hiekkalinna T, Leppavuori J, Pajukanta P, Wessman M, Koskenvuo M, Palotie A, Lange K, Kaprio J, and Peltonen L (2001) Quantitative-trait-locus analysis of body-mass index and of stature, by combined analysis of genome scans of five Finnish study groups. *Am J Hum Genet* 69:117-123

Perou CM (2001) Show me the data! *Nat Genet* 29:373

Pratt SC, Daly MJ, and Kruglyak L (2000) Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *Am J Hum Genet* 66:1153-1157

Preece MA (1996) The genetic contribution to stature. *Horm Res* 45 Suppl 2:56-58

Province MA (2001) The significance of not finding a gene. *Am J Hum Genet* 69:660-663

Raivio T, Huhtaniemi I, Anttila R, Siimes MA, Hagenas L, Nilsson C, Pettersson K, and Dunkel L (1996) The role of luteinizing hormone-beta gene polymorphism in the onset and progression of puberty in healthy boys. *J Clin Endocrinol Metab* 81:3278-3282

Rao E, Weiss B, Fukami M et al (1997) Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat Genet* 16:54-63

Rappold GA, Fukami M, Niesler B, Schiller S, Zumkeller W, Bettendorf M, Heinrich U, Vlachopapadoupoulou E, Reinehr T, Onigata K, and Ogata T (2002) Deletions of the homeobox gene SHOX (short stature homeobox) are an important cause of growth failure in children with short stature. *J Clin Endocrinol Metab* 87:1402-1406

Rijsdijk FV and Sham PC (2002) Estimation of sib-pair IBD sharing and multipoint polymorphism information content by linear regression. *Behav Genet* 32:211-220

Risch N (1990a) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet* 46:229-241

Risch N (1990b) Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet* 46:242-253

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

Rosenthal R and Rubin DB (2003) r equivalent: A simple effect size indicator. *Psychol Methods* 8:492-496

S.A.G.E. (2004) Statistical Analysis for Genetic Epidemiology. *Release 5 0*

Sagoo GS, Tazi-Ahnini R, Barker JW, Elder JT, Nair RP, Samuelsson L, Traupe H, Trembath RC, Robinson DA, and Iles MM (2004) Meta-analysis of genome-wide studies of psoriasis susceptibility reveals linkage to chromosomes 6p21 and 4q28-q31 in Caucasian and Chinese Hans population. *J Invest Dermatol* 122:1401-1405

Sale MM, Freedman BI, Hicks PJ, Williams AH, Langefeld CD, Gallagher CJ, Bowden DW, and Rich SS (2005) Loci Contributing to Adult Height and Body Mass Index in African American Families Ascertained for Type 2 Diabetes. *Ann Hum Genet* 69:1-11

Sammalisto SA, Hiekkalinna T, Suviolahti E et al (2005) A male-specific QTL on 1p21 controlling human stature. *J Med Genet* 42:932-939

Schneider KU, Sabherwal N, Jantz K, Roth R, Muncke N, Blum WF, Cutler GB, Jr., and Rappold G (2005) Identification of a major recombination hotspot in patients with short stature and SHOX deficiency. *Am J Hum Genet* 77:89-96

Schork NJ and Greenwood TA (2004) Inherent bias toward the null hypothesis in conventional multipoint nonparametric linkage analysis. *Am J Hum Genet* 74:306-316

Schubbert S, Zenker M, Rowe SL et al (2006) Germline KRAS mutations cause Noonan syndrome. *Nat Genet* 38:331-336

Schuit SC, van Meurs JB, Bergink AP, van der KM, Fang Y, Leusink G, Hofman A, van Leeuwen JP, Uitterlinden AG, and Pols HA (2004) Height in pre- and postmenopausal women is influenced by estrogen receptor alpha gene polymorphisms. *J Clin Endocrinol Metab* 89:303-309

Self S and Liang K-Y (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J Am Stat Assoc* 82:605-610

Sham PC. 1998. Statistics in Human Genetics. London: Arnold

Sham PC, Purcell S, Cherny SS, and Abecasis GR (2002) Powerful regression-based quantitative-trait linkage analysis of general pedigrees. *Am J Hum Genet* 71:238-253

Sieberts SK, Broman KW, and Gudbjartsson DF (2004) "Bias toward the null" means reduced power. *Am J Hum Genet* 75:720-722

Silventoinen K (2003) Determinants of variation in adult body height. *J Biosoc Sci* 35:263-285

Silventoinen K, Kaprio J, Lahelma E, and Koskenvuo M (2000) Relative effect of genetic and environmental factors on body height: differences across birth cohorts among Finnish men and women. *Am J Public Health* 90:627-630

Simpson EH (1951) The interpretation of interaction in contingency tables. *Journal of the American Statistical Association* 88:81-88

Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams RM. 1949. The American Soldier: Adjustment During Army Life. Vol. 1. New Jersey: Princeton University Press

Stuppia L, Calabrese G, Gatta V et al (2003) SHOX mutations detected by FISH and direct sequencing in patients with short stature. *J Med Genet* 40:E11

Suarez F, Rossignol C, and Garabedian M (1998) Interactive effect of estradiol and vitamin D receptor gene polymorphisms as a possible determinant of growth in male and female infants. *J Clin Endocrinol Metab* 83:3563-3568

Suarez F, Zeghoud F, Rossignol C, Walrant O, and Garabedian M (1997) Association between vitamin D receptor gene polymorphism and sex-dependent growth during the first two years of life. *J Clin Endocrinol Metab* 82:2966-2970

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. 2000. Methods for Meta-Analysis in Medical Research. Chichester: John Wiley & Sons

Tang HK and Siegmund D (2001) Mapping quantitative trait loci in oligogenic models. *Biostatistics* 2:147-162

Tao C, Yu T, Garnett S, Briody J, Knight J, Woodhead H, and Cowell CT (1998) Vitamin D receptor alleles predict growth and bone density in girls. *Arch Dis Child* 79:488-493

The ADHD Molecular Genetics Network (2000) Collaborative possibilities for molecular genetic studies of attention deficit hyperactivity disorder: report from an international conference. *Am J Med Genet* 96:251-257

The Transatlantic Multiple Sclerosis Genetics Cooperative (2001) A meta-analysis of genomic screens in multiple sclerosis. *Mult Scler* 7:3-11

Thompson DB, Ossowski V, Janssen RC, Knowler WC, and Bogardus C (1995) Linkage between stature and a region on chromosome 20 and analysis of a candidate gene, bone morphogenetic protein 2. *Am J Med Genet* 59:495-500

Tippett L.H.C. 1931. The method of statistics. London: Williams & Norgate

van der Sluis I, de Muinck Keizer-Schrama SM, Krenning EP, Pols HA, and Uitterlinden AG (2003) Vitamin D receptor gene polymorphism predicts height and bone size, rather than bone density in children and young adults. *Calcif Tissue Int* 73:332-338

Van Eerdewegh P, Little RD, Dupuis J et al (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* 418:426-430

Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. *Science* 291:1304-1351

Visscher PM and Wray NR (2004) Conventional multipoint nonparametric linkage analysis is not necessarily inherently biased. *Am J Hum Genet* 75:718-720

Willemsen G, Boomsma DI, Beem AL, Vink JM, Slagboom PE, and Posthuma D (2004) QTLs for height: results of a full genome scan in Dutch sibling pairs. *Eur J Hum Genet* 12:820-828

Wiltshire S, Frayling TM, Hattersley AT, Hitman GA, Walker M, Levy JC, O'Rahilly S, Groves CJ, Menzel S, Cardon LR, and McCarthy MI (2002)

Evidence for linkage of stature to chromosome 3p26 in a large U.K. Family data set ascertained for type 2 diabetes. *Am J Hum Genet* 70:543-546

Wise LH (2001) Inclusion of candidate region studies in meta-analysis using the genome screen meta-analysis method: application to asthma data. *Genet Epidemiol* 21 Suppl 1:S160-S165

Wise LH, Lanchbury JS, and Lewis CM (1999) Meta-analysis of genome searches. *Ann Hum Genet* 63:263-272

Wjst M, Fischer G, Immervoll T et al (1999) A genome-wide search for linkage to asthma. *Genomics* 58:1-8

Wu X, Cooper RS, Boerwinkle E, Turner ST, Hunt S, Myers R, Olshen RA, Curb D, Zhu X, Kan D, and Luke A (2003) Combined analysis of genomewide scans for adult height: results from the NHLBI Family Blood Pressure Program. *Eur J Hum Genet* 11:271-274

Wu X, Cooper RS, Borecki I, Hanis C, Bray M, Lewis CE, Zhu X, Kan D, Luke A, and Curb D (2002) A combined analysis of genomewide linkage scans for body mass index from the national heart, lung, and blood institute family blood pressure program. *Am J Hum Genet* 70:1247-1256

Xu J, Bleecker ER, Jongepier H, Howard TD, Koppelman GH, Postma DS, and Meyers DA (2002) Major Recessive Gene(s) with Considerable Residual Polygenic Effect Regulating Adult Height: Confirmation of Genomewide Scan Results for Chromosomes 6, 9, and 12. *Am J Hum Genet* 71:646-650

Xu J, Meyers DA, Ober C et al (2001) Genomewide screen and identification of gene-gene interactions for asthma-susceptibility loci in three U.S. populations: collaborative study on the genetics of asthma. *Am J Hum Genet* 68:1437-1446

Zaykin DV, Zhivotovsky LA, Westfall PH, and Weir BS (2002) Truncated product method for combining P-values. *Genet Epidemiol* 22:170-185

Zintzaras E and Ioannidis JP (2005) Heterogeneity testing in meta-analysis of genome searches. *Genet Epidemiol* 28:123-137

Zmuda JM, Cauley JA, Kuller LH, and Ferrell RE (2001) A common promotor variant in the cytochrome P450c17alpha (CYP17) gene is associated with bioavailability testosterone levels and bone size in men. *J Bone Miner Res* 16:911-917

11 Lebenslauf

Astrid Dempfle
Geb. am 14.1.1975 in Freising
Staatsangehörigkeit: deutsch

Sept. 1981 - Juli 1985	Grundschule Kranzberg
Sept. 1985 - Juli 1994	Josef-Hofmiller-Gymnasium, Freising
Aug. 1991 - Juli 1992	High School Besuch in Sioux City, Iowa, USA
Juli 1994	Abitur am Josef-Hofmiller-Gymnasium, Freising
Okt. 1994 - Aug. 1996	Mathematikstudium an der Justus-Liebig-Universität Gießen mit Nebenfach Stochastik
Sept. 1996 - Feb. 1997	Mathematikstudium an der Université des Sciences et Techniques du Languedoc, Montpellier, Frankreich
März 1997 - März 2000	Fortsetzung des Mathematikstudiums an der Justus-Liebig-Universität Gießen Diplomarbeit bei Prof. Stute: "Changepoints in der Regression"
27. März 2000	Diplom in Mathematik (Note: sehr gut)
Mai 2000 - April 2001	Wissenschaftliche Mitarbeiterin am Institut für Epidemiologie (AG Genetische Epidemiologie) der GSF, München
Mai 2001 - Feb 2004	Wissenschaftliche Mitarbeiterin am Institut für Medizinische Biometrie und Epidemiologie, Philipps-Universität Marburg
März 2004 - August 2004	Familienpause (eine Tocher, geb. 31.3.2004)
Seit Sept. 2004	Wissenschaftliche Mitarbeiterin am Institut für Medizinische Biometrie und Epidemiologie, Philipps-Universität Marburg

Eigene Publikationen

1. Loesgen S, Dempfle A, Golla A, Bickeböller H (2001). Weighting Schemes in Pooled Linkage Analysis. *Genet. Epidemiol*. 21 Suppl 1: S142-S147
2. Immervoll Th, Loesgen S, Dütsch G, Gohlke H, Herbon N, Klugbauer S, Dempfle A, Bickeböller H, Becker-Follmann J, Rüschendorf F, Saar K, Reis A, Wichmann HE, Wjst M (2001). Fine mapping and single nucleotide polymorphism association results of candidate genes for asthma and related phenotypes. *Human Mutation* 18: 327-336.
3. Dempfle A, Stute W (2002). Nonparametric Estimation of a Discontinuity in Regression. *Statictica Neerlandica*. 56: 233-242
4. Scherag A, Dempfle A, Hinney A, Hebebrand J, Schäfer H (2002). Confidence intervals for genotype relative risks in case-parent trio designs for candidate gene association studies. *Human Heredity*. 54:210-217
5. Hebebrand J, Wulftange H, Dempfle A, Geller F, Hinney A (2002). Epidemic obesity: are genetic factors involved via increased rates of assortative mating? *Obesity*. 26 Suppl 1:S67
6. Smidt J, Heiser P, Dempfle A, Konrad K, Hemminger U, Kathöfer A, Halbach A, Strub J, Grabarkiewicz J, Kiefl H, Linder M, Knölker U, Warnke A, Remschmidt H, Herpertz-Dahlmann B, Hebebrand J (2003). Formalgenetische Befunde zur Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung. *Fortschritte der Neurologie · Psychiatrie*. 71:366-377
7. Geller F*, Dempfle A*, Görg T (2003). Genome Scan for BMI and height in the Framingham Heart Study. *BMC Genetics*. 4 Suppl 1:S91 (* gemeinsame Erstautorenschaft)
8. Scherag A, Müller HH, Dempfle A, Hebebrand J, Schäfer H (2003). Data adaptive interim recalculation of sample sizes for candidate-gene association studies. *Human Heredity*. 56:56-62
9. Fritsch B, Karakizlis H, Hamer HM, Dempfle A, Oertel WH, Rosenow F (2003). Long-term effects of chronic-intermittent deep brain stimulation of the ncl. subthalamicus (STN) on seizure activity in a rat model of absence epilepsy. *Epilepsia.* 44 Suppl 9:63-64
10. Dempfle A*, Loesgen S* (2004). Meta-analysis of linkage studies for complex diseases: an overview of methods and a simulation study. *Annals of Human Genetics*. 68:69-83 (* gemeinsame Erstautorenschaft)

11. Malik S, Arshad M, Din MA-U, Oeffner F, Dempfle A, Haque S, Koch M, Ahmad W, Grzeschik K-H (2004). A novel type of autosomal recessive syndactyly: clinical and molecular studies in a family of Pakistani origin. *Am J Med Genet*. 126A:61-77
12. Wang HJ, Geller F, Dempfle A, Schäuble N, Friedel S, Lichtner P, Fontenla-Horro F, Wudy S, Hagemann S, Gortner L, Huse K, Remschmidt H, Bettecken T, Meitinger T, Schäfer H, Hebebrand J, Hinney A (2004). Ghrelin Receptor Gene: Identification of several sequence variants in extremely obese children and adolescents, healthy normal weight and underweight students and children with short normal stature. *J Clin Endocrinol Metab*. 89:157-162
13. Geller F, Reichwald K, Dempfle A, Illig T, Vollmert C, Herpertz S, Siffert W, Platzer M, Hess C, Gudermann T, Biebermann H, Wichmann HE, Schäfer H, Hinney A, Hebebrand J (2004). Melanocortin-4 receptor gene variant I103 is negatively associated with obesity. *Am J Hum Genet*. 74:572-581
14. Hebebrand J, Geller F, Dempfle A, Heinzel-Gutenbrunner M, Raab M, Gerber G, Wermter AK, Fontenla Horro F, Blundell J, Schäfer H, Remschmidt H, Herpertz S, Hinney A (2004). Binge-eating episodes are not characteristic of carriers of melanocortin-4 receptor gene mutations. *Mol Psychiatry*. 9:796-800
15. Sure U, Battenberg E, Dempfle A, Tirakotai W, Bien S, Bertalanffy H (2004). Local Hypoxia Upregulates Endothelial Proliferation and Neoangiogenesis in Cerebral Arteriovenous Malformations. *Neurosurgery*. 55:663-670
16. Dempfle A, Hinney A, Heinzel-Gutenbrunner M, Raab M, Geller F, Gudermann T, Schäfer H, Hebebrand J (2004). Large Quantitative Effect of Melanocortin-4 Receptor Gene Mutations on BMI. *J Med Genet*. 41:795-800
17. Heiser P, Friedel S, Dempfle A, Konrad K, Smidt J, Grabarkiewicz J, Herpertz-Dahlmann B, Remschmidt H, Hebebrand J (2004). Molecular Genetic Aspects of Attention-Deficit/Hyperactivity Disorder. *Neuroscience and Biobehavioural Reviews*. 28:625-641
18. Friedel S, Fontenla Horro F, Wermter AK, Geller F, Dempfle A, Reichwald K, Smidt J, Konrad K, Herpertz-Dahlmann B, Warnke A, Hemminger U, Linder M, Kiefl H, Goldschmidt HP, Siegfried W, Remschmidt H, Hinney A, Hebebrand J (2004). Mutation screen of the brain derived neurotrophic factor gene (BDNF): Identification of several genetic variants and association studies

in patients with obesity, eating disorders and attention-deficit/hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet*. 132B:96-99

19. Seifart C, Dempfle A, Plagens A, Seifart U, Clostermann U, Müller B, Vogelmeier C, von Wichert P (2005). TNF-α, TNF-β, IL-6 and IL-10 promotor polymorphisms in patients with chronic obstructive pulmonary disease. *Tissue Antigens.* 65:93-100
20. Wudy SA, Hagemann S, Dempfle A, Ringler G, Blum WF, Berthold LD, Alzen G, Gortner L, Hebebrand J (2005). Children with Idiopathic Short Stature are Poor Eaters and Have Decreased Body Mass Index. *Pediatrics.*116: e52-e57
21. Wündisch T, Thiede C, Morgner A, Dempfle A, Günther A, Liu H, Ye H, Du MQ, Kim TD, Du M, Bayerdörffer E, Stolte M, Neubauer A (2005). Long-term follow up of patients with gastric MALT lymphoma after *Helicobacter pylori* eradication. *J Clin Oncol*. 23:8018-8024
22. Seifart C, Plagens A*, Dempfle A*, Clostermann U, Vogelmeier C, von Wichert P, Seifart U (2005). TNF-α, TNF-β, IL-6 and IL-10 polymorphisms in patients with lung cancer. *Disease Markers*. 21:157-165 (* gemeinsame Zweitautorenschaft)
23. Walitza S*, Renner T*, Dempfle A*, Wewetzer C, Halbach A, Herpertz-Dahlmann B, Konrad K, Remschmidt H, Linder M, Schäfer H, Gross C, Hebebrand J, Warnke A, Lesch KP (2005). Transmission disequilibrium of polymorphic variants in the tryptophan hydroxylase-2 gene in attention-deficit/hyperactivity disorder. *Mol Psychiatry.* 10:1126-1132 (* gemeinsame Erstautorenschaft)
24. Hebebrand J, Dempfle A, Saar K, Thiele H, Herpertz-Dahlmann B, Linder M, Kiefl H, Remschmidt H, Hemminger U, Warnke A, Knölker U, Heiser P, Friedel S, Hinney A, Schäfer H, Nürnberg P, Konrad K (2006). A Genome-wide Scan for Attention Deficit/Hyperactivity Disorder in 155 German Sib Pairs. *Mol Psychiatry.* 11:196-205
25. Renner T, Walitza S, Dempfle A, Konrad K, Wewetzer C, Halbach A, Herpertz-Dahlmann B, Remschmidt H, Smidt J, Linder M, Flierl L, Knolker U, Friedel S, Schafer H, Gross C, Hebebrand J, Warnke A, Lesch KP (2006). Polymorphic variants in the transcriptional control region of TPH2 are

preferentially transmitted in ADHD. *European Neuropsychopharmacology*. 16(Suppl. 1): S21-S22

26. Friedel S, Heiser P, Dempfle A, Konrad K, Hebebrand J (2006). Molecular Genetic Aspects of Attention-Deficit/Hyperactivity Disorder. In: Attention-Deficit/Hyperactivity Disorder (AD/HD) and the Hyperkinetic Syndrome (HKS): Current Ideas and Ways Forward. Ed. Robert D. Oades. Nova Science Publishers, Hauppauge, NY
27. Zeiler T, Heim M, Dempfle A, Kretschmer V (2006). Platelets do not adsorb HLA class I molecules during storage of pooled platelet concentrates. *Transfus Med*. 16:176-183
28. Dempfle A, Wudy SA, Saar K, Hagemann S, Friedel S, Scherag A, Berthold LD, Alzen G, Gortner L, Blum WF, Hinney A, Nürnberg P, Schäfer H, Hebebrand J (2006). Evidence for involvement of the vitamin D receptor gene in idiopathic short stature via a genome wide linkage study and subsequent association studies. *Human Molecular Genetics*. 15(18):2772-2783
29. Görg Ch, Bert T, Kring R, Dempfle A (2006). Transcutaneous contrast enhanced sonography of the chest for evaluation of pleural based pulmonary lesions: experience in 137 patients. *Ultraschall Med*. 27(5):437-444
30. Schimmelmann BG, Friedel S, Christiansen H, Dempfle A, Hinney A, Hebebrand J. Genetische Befunde bei der Aufmerksamkeitsdefizit- und Hyperaktivitätsstörung (ADHS) (2006). *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie.* 34(6):425-433
31. Heiser P, Dempfle A, Friedel S, Konrad K, Hinney A, Kiefl H, Walitza S, Bettecken T, Saar K, Linder M, Warnke A, Herpertz-Dahlmann B, Schäfer H, Remschmidt H, Hebebrand J. Family-based association study of serotonergic candidate genes and attention-deficit/hyperactivity disorder in a German sample. *J Neural Transmission.* In press
32. Schimmelmann BG, Friedel S, Dempfle A, Warnke A, Lesch KP, Walitza S, Renner T, Romanos M, Herpertz-Dahlmann B, Linder M, Schäfer H, Seitz C, Palmason H, Frei-tag C, Meyer J, Konrad K, Hinney A, Hebebrand J. No Evidence for Preferential Transmission of Common Valine66 Allele of the Val66Met Polymorphism of the Brain-Derived Neurotrophic Factor Gene (BDNF) In ADHD. *J Neural Transmission.* In press

12 Akademische Lehrer

Meine akademischen Lehrer waren die Damen und Herren

in Gießen
Bartsch, Beutelspacher, Braunss, Fenske, Häusler, Hübner, Metsch, Stute, Timmesfeld, Walther

in Montpellier
Attouch, Boyom, Ducharme

in München
Bickeböller, Wichmann

in Marburg
Grzeschik, Hebebrand, Schäfer.

13 Danksagung

Zunächst danke ich Prof. Dr. Helmut Schäfer für die stete Unterstützung bei der Erstellung dieser Doktorarbeit und für die Möglichkeit in den vergangenen Jahren an einer Vielzahl sehr interessanter und spannender Projekte zu arbeiten. Prof. Dr. Heike Bickeböller danke ich für die gute Betreuung bei den ersten Schritten in die Genetische Epidemiologie.

Besonders danke ich Sabine Loesgen für ihre Unterstützung bei der Einarbeitung in dieses Thema. Sie trug mit etlichen Ideen und konstruktiven Anregungen zum Gelingen der Arbeit bei, insbesondere war sie an der Durchführung der Asthma Meta-Analyse und der Simulationsstudie beteiligt. Frank Geller danke ich für die sehr gute Zusammenarbeit in vielen Projekten, speziell dem Genetic Analysis Workshop 13 und ganz besonders für das aufmerksame Korrekturlesen dieser Arbeit. Gundula Huth und Tilman Görg verdienen großen Dank für ihre Mitarbeit an der Meta-Analyse zur Körpergröße, am Genetic Analysis Workshop 13 und für all die vielen Aufgaben, die sie so sorgfältig und gewissenhaft erledigen. Auch André Scherag gebührt besonderer Dank für viele anregende Diskussionen und seine Hilfsbereitschaft. Allen derzeitigen und ehemaligen Kollegen danke ich für viele kleine und große Hilfen, gute Tipps, interessante Gespräche und insbesondere für das sehr gute Arbeitsklima am Institut für Medizinische Biometrie und Epidemiologie in Marburg.

Auch von etlichen guten Kooperationen mit klinischen und molekulargenetischen Partnern habe ich profitiert. Besonders bereichernd ist die Zusammenarbeit mit Prof. Dr. Johannes Hebebrand und seinen Mitarbeitern, insbesondere Susann Friedel und Dr. Anke Hinney, denen ich für die Gelegenheit danke, in gemeinsamen Projekten so viel zu lernen.

Ich danke den Organisatoren des Genetic Analysis Workshop und den Wissenschaftlern, die für diese Workshops ihre Daten (Asthma Genomscans und

Framingham Heart Study) zur Verfügung gestellt haben. Ebenso danke ich den Wissenschaftlern, die mir persönlich ihre detaillierten Ergebnisse für die Meta-Analyse zur Körpergröße zur Verfügung gestellt haben.

Ich danke meinen Eltern dafür, dass sie mich nie gedrängt sondern immer unterstützt und ermutigt haben und dafür, dass sie stets viel Interesse an meiner Arbeit zeigen. Allen Freunden und Verwandten danke ich besonders für die Ablenkungen von der Arbeit und gemeinsam verbrachte Stunden. Ina Lochnit danke ich für die beste Kinderbetreuung, die man sich vorstellen kann, die es mir ermöglicht immer unbeschwert zu arbeiten. Julie danke ich für die Freude und den Spaß, den sie in unser Leben bringt. Und schließlich bin ich meinem Lebensgefährten Tobias Donath besonders dankbar für seine fortwährende Unterstützung, ohne die ich diese Arbeit nicht hätte fertig stellen können.

14 Ehrenwörtliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die dem Fachbereich Medizin der Philipps-Universität Marburg zur Promotionsprüfung eingereichte Arbeit mit dem Titel „Evaluation of methods for meta-analysis of genetic linkage studies for complex diseases and application to genome scans for asthma and adult height“ am Institut für Medizinische Biometrie und Epidemiologie unter der Leitung von Prof. Dr. H. Schäfer mit Unterstützung durch Frau Prof. Dr. H. Bickeböller und Frau Dipl.-Stat. S. Loesgen (damals Institut für Epidemiologie der GSF, München) ohne sonstige Hilfe selbst durchgeführt und bei der Abfassung der Arbeit keine anderen als die in der Dissertation angeführten Hilfsmittel benutzt habe.

Ich habe bisher an keinem in- und ausländischen medizinischen Fachbereich ein Gesuch um Zulassung zur Promotion eingereicht noch die vorliegende oder eine andere Arbeit als Dissertation vorgelegt.

Teile der vorliegenden Arbeit wurden in den Zeitschriften *Genetic Epidemiology* (2001, 21 Supplement 1: S142-S147), *BMC Genetics* (2003, 4 Supplement 1:S91) und *Annals of Human Genetics* (2004, 68:69-83) veröffentlicht.
Außerdem wurden Teile der vorliegenden Arbeit als Poster oder Vortrag auf folgenden Tagungen vorgestellt: *Genetic Analysis Workshop 12*, 23.-26.10.2000, San Antonio, Texas (USA); *47. Biometrisches Kolloquium*, 20.-23.3.2001, Homburg/Saar; *European Mathematical Genetics Meeting*, 6.-8.4.2002, Sheffield, Großbritannien; Symposium *„Genomics of Chronic Inflammatory Diseases“* des NGFN, 5.-7.7.2002, Kiel; *Genetic Analysis Workshop 13*, 11.-14.11.2002, New Orleans, Louisiana (USA); *11. International Genetic Epidemiology Society Meeting*, 15.-16.11.2002, New Orleans, Louisiana (USA); NGFN Partnering-Day *„Disorders of Body Weight Regulation“*, 29.-31.1.2003, Marburg; Tagung *„Medical Genetics - Current Developments in Statistical Methodology for Genetic Architecture of Complex Diseases“*, 2.2.-8.2.2003, Mathematisches Forschungsinstitut Oberwolfach; *14. Jahrestagung der Deutschen Gesellschaft für Humangenetik*, 1.-4.10.2003, Marburg.

Marburg, den 30.7.2006