Compare and Contrast Meta Analysis (CCMA): An Application for Genomewide Association Studies

Hansjörg Baurecht

Dissertation an der
Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität

München, den 13. August 2016

Compare and Contrast Meta Analysis (CCMA): An Application For Genomewide Association Studies

Dissertation an der
Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität
zur Erlangung des Doktorgrades (Dr. rer. nat.)

vorgelegt von
Hansjörg Baurecht
am 13. August 2016

Erstgutachter: Prof. Dr. Konstantin Strauch
Zweitgutacher: Prof. Dr. Thomas Augustin
Drittgutachter: Prof. Dr. Michael Nothnagel
Tag der Disputation: 27. Januar 2017

"Was lange währt, wird endlich gut."

(Deutsches Sprichwort)

Meiner Oma

List of Abbreviations

AE atopic eczema

ASSET R package, which implemented the Subset-Based Meta Analysis

C11orf30 chromosome 11 Open Reading Frame 30

CCMA Compare & Contrast Meta Analysis

CD-CV common diseases–common variant

CI confidence interval

COM COMBINED & OVERLAP method

DL DerSimonian & Laird's method for estimating the between-study variance

DLM discrete local maxima method

DNA deoxyribonucleic acid

EAGLE EArly Genetics and Lifecourse Epidemiology consortium

FCER1A Fc fragment of IgE, high affinity I receptor alpha

FLG filaggrin

GWAS genome wide association studies

HBB hemoglobin subunit beta gene

HLA Human Leukocyte Antigen

HWE Hardy-Weinberg equilibrium

LD linkage disequilibrium

LRRC32 leucine rich repeat containing 32

MAF minor allele frequency

MCMC Markov chain Monte Carlo

ML maximum likelihood estimate

MNM multinomial regression model

MOM methods of moment

OR odds ratio

RAD50 RAD50 double strand break repair protein

REML restricted maximum likelihood estimate

SBMA Subset-Based Meta Analysis

SNP single nucleotide polymorphism

SNV single nucleotide variant

TDT transmission disequilibrium test

wCCMA weighted Compare & Contrast Meta Analysis

Contents

Summary

Genetic epidemiology has moved from linkage studies via candidate gene association studies towards genome wide association studies (GWAS), which use single nucleotide variants (SNVs), the smallest genetic entity, to map diseases to susceptibility loci in a hypothesis free way. The era of GWAS started in 2005, when high throughput genotyping became feasible and affordable, and it has since been a great success. GWAS have revealed many markers associated with complex diseases, and meta-analyses of several GWAS further increased the set of known disease-related variants. Besides the identification of disease specific effects, GWAS-based methods revealed susceptibility loci shared between several clinically related and also unrelated diseases. These pleiotropic loci are of high biological interest considering that they may mark shared or branching pathophysiological mechanisms.

Customized genotyping arrays, e.g. the Immunochip, further supported the view that several diseases map to the same susceptibility loci showing agonistic and antagonistic effects. Some methods have been proposed for investigating and searching for genetic overlap between diseases, however there is still a need for sound statistical methods to address this issue.

The aim of this thesis was to develop a method based on meta-analysis techniques to compare and contrast two complex diseases, in particular, to find agonistic and antagonistic loci that may contribute to the understanding of the genetic architecture of psoriasis and atopic eczema.

This work presents a new method, the **C**ompare & **C**ontrast **M**eta **A**nalysis (CCMA), which allows researchers to compare and contrast two diseases on a genetic basis. Meta-analysis techniques are used to impose a test statistic that allows for identifying agonistic and antagonistic effects. A closed form for the density and cumulative distribution function of the CCMA test statistic is presented, which, conveniently for practical reasons, turns out to be exponentially distributed. Hence, thresholds for suggestive and genome-wide significant association can easily be derived and, in contrast to the already reported Subset-Based Meta Analysis (SBMA), the mode of pleiotropy can be inferred directly. Modified versions of the test statistic allow incorporating study size, which, depending on the transformation matrix, improve the power for detecting agonistic or antagonistic effects.

The power and type I error of the CCMA method are compared with those of the SBMA method by simulation. The CCMA method shows marginally lower power than SBMA but the type I error is better controlled.

The CCMA method is applied to several published GWAS on atopic eczema and psoriasis. In order to compare it with the computer intensive multinomial regression models (MNM), the SBMA is used to reduce the vast amount of SNVs to those showing at least suggestive

disease-specific or pleiotropic effects. These SNVs are reanalyzed using MNM, and their effect categorization (disease-specific or pleiotropic) is compared with the ones obtained by the CCMA method. The comparison reveals high agreement of 85.5% overall and 93.6% without the most complex Human Leukocyte Antigen (HLA) region, which demonstrates the usability of the new method. Finally, the CCMA is compared with the COMPARED & OVERLAP approach in terms of identification of associated SNVs and concordance of effect categorization with the MNM as gold standard.

This work shows that the CCMA method is an appealing approach to identify disease-specific and pleiotropic loci using available GWAS data and effectively exploits additional cross-phenotype information. It shows comparable power to the SBMA method while better controlling the type 1 error and outperforms the COMPARED & OVERLAP approach. It shows high agreement with the MNM in terms of effect categorization. Finally, the CCMA method can also be applied to other genome-wide molecular data such as gene expression, epigenomics or metabolomics, as well as to other research questions that arise in environmental epidemiology. In that context, the influence of lifestyle factors or environmental exposures on two different diseases can be investigated with regard to their concordant or contrasting effect.

Zusammenfassung

Die genetische Epidemiologie hat sich in den letzten Jahren von Kopplungsstudien über Kandidatengen-Assoziationsstudien hin zu genomweiten Assoziationsstudien (GWAS) entwickelt, welche basierend auf der kleinsten genetischen Einheit, den Einzelnukleotidvarianten (single nucleotide variants, SNVs), Erkrankungen auf agnostische Weise im Genom kartiert. Die Ära der GWAS, die 2005 begann, als die Hochdurchsatzgenotypisierung zu annehmbaren Kosten technisch realisierbar wurde, war ein großer Erfolg. GWAS haben viele Marker, die mit komplexen Erkrankungen assoziiert sind, identifiziert, und Meta-Analysen von mehreren GWAS zur gleichen Erkrankung haben die Menge der bekannten krankheitsbezogenen Varianten erweitert. Neben krankheitsspezifischen Effekten wurden mit GWAS-basierten Methoden gemeinsame Loci von klinisch verwandten und nicht verwandten Erkrankungen entdeckt. Diese pleiotropen Loci sind von großem biologischem Interesse, da sie gemeinsame und verzweigende pathophysiologische Mechanismen anzeigen.
Die erfolgreiche Verwendung nutzerspezifischer Genchips, wie z.B. des Immunochip, belegen die Vermutung, dass verschiedenen Erkrankungen dieselben Suszeptibilitätsloci zugrunde liegen, die agonistische oder antagonistische Effekte aufweisen. Es wurden bereits statistische Methoden entwickelt, um gemeinsame genetische Komponenten verschiedener Erkrankungen zu identifizieren, jedoch besteht weiterhin Forschungsbedarf in diesem Bereich.
Ziel dieser Arbeit ist es, eine auf Meta-Analyse basierende Methode zu entwickeln, um zwei komplexe Erkrankungen zu vergleichen und zu kontrastieren, insbesondere um agonistische und antagonistische Loci zu identifizieren, die zum Verständnis der genetischen Grundlage der Psoriasis und des atopischem Ekzems beitragen.
Die vorliegende Arbeit präsentiert die **C**ompare & **C**ontrast **M**eta **A**nalysis (CCMA) Methode, die es erlaubt, zwei Erkrankungen hinsichtlich ihrer genetischen Grundlage zu vergleichen und zu kontrastieren. Dazu wird eine Teststatistik basierend auf Meta-Analyse-Ergebnissen entwickelt, die agonistische und antagonistische Effekte untersucht. Für die Verteilung der Teststatistik kann eine geschlossene Form angegeben werden, die kritische Werte für suggestive und genomweite signifikante Assoziationen liefert. Analysen zur Power und zum Fehler 1. Art werden durchgeführt, um die CCMA-Methode mit der bereits bekannten Subset-Based Meta-Analyse (SBMA) zu vergleichen. Die CCMA-Methode wird auf verschiedene GWAS-Daten zur Psoriasis und zum atopischen Ekzem angewendet und selektiert SNVs mit krankheitsspezifischen oder pleiotropen Effekten, die die Signifikanzschranke für suggestive Assoziation unterschreiten. Diese SNVs werden mit Hilfe des multinomialen Regressionsmodells (MNM) erneut analysiert und die Kategorisierung in krankheitsspezifische oder pleiotrope Effekte mit jener der CCMA-Methode verglichen. Eine große Übereinstimmung beider Methoden von 85.5% insgesamt

bzw. 93.6% ohne die komplexe humane Leukozytenantigenregion (HLA-Region) bestätigt die Verwendbarkeit der neuen Methode. Schließlich wird die CCMA mit der SBMA und der COMPARED & OVERLAP Methode hinsichtlich identifizierter SNVs und Übereinstimmung der Effektkategorisierung mit dem MNM als Goldstandard verglichen.
Die CCMA-Methode basiert auf Meta-Analyse-Teststatistiken mit guter Power und ist schnell und einfach zu implementieren. Sie kann sowohl pleiotrope als auch krankheitsspezifische Marker identifizieren. Im Vergleich zu anderen Methoden liegen die Vorteile der CCMA in ihrer Einfachheit und Präzision, ohne individuelle Genotypdaten verwenden zu müssen. Die hohe Übereinstimmung mit dem multinomialen Regressionsmodell als Goldstandard bei der Effektkategorisierung unterstreicht die Nutzbarkeit der neuen Methode.

1. Introduction

In recent years, the field of genetic epidemiology has experienced a rapid improvement in genotyping technology. The focus to study genetic variability moved from linkage studies and candidate gene association approaches to genomewide association studies (GWAS) as high throughput genotyping of vast amounts of single nucleotide variants (SNVs) at lower costs became feasible. With the advent of next-generation sequencing a new era of studying genetic variability approached, which allows researchers to focus on rare exonic and non-coding variants. These developments are followed by changes in data processing and statistical methods to meet the requirements to handle enormous data volume and to adequately analyze them.

GWAS have been very successful in recent years to unravel new susceptibility loci responsible for complex diseases and have identified many loci shared among common disorders.[1] Customized arrays have been designed by consortia of related diseases, e.g. the Immunochip for immune-mediated disorders, to fine map established GWAS loci at high resolution and to identify SNVs shared among different traits. With the Immunochip, new risk loci have been identified for psoriasis[2] and atopic eczema (AE),[3] and both agonistic and antagonistic effects were observed for these and other immune-mediated diseases.

The interest in pleiotropy, "the multi-functionality of a gene in phenotype presentation",[4] increased in recent years, and several methods[1,4] have been proposed for an appropriate analysis using external sources such as the GWAS catalog.[5] Other approaches exploit GWAS results using meta-analysis based methods.[6,7]

First we give an overview of the developments in genetic epidemiology in general, before we draw attention to the investigation of pleiotropic loci.

1.1. Basics of Genetics

The human genome consists of a diploid set of 23 chromosomes with 22 pairs of homologous autosomes and a pair of sex chromosomes. Females possess two X-chromosomes whereas males possess one X- and one Y-chromosome.
A gene is defined as "a union of genomic sequences encoding a coherent set of potentially overlapping functional products".[8] The position of a defined DNA sequence – coding or non-coding – on a human chromosome is called a (gene) locus whereas different variants of this sequence are termed alleles. Mutations are changes of the DNA and may occur spontaneously (e.g. error in DNA replication) or may be triggered by extrinsic sources, e.g. ionizing radiation etc. Single nucleotide variants (SNVs) are the smallest entity of genetic variation and are defined as an exchange of a nucleotide, e.g. the nucleotide A is replaced by C. If the SNVs exhibit minor allele frequencies >1%, they are often referred to as common variants and are synonymously called single nucleotide polymorphisms (SNPs).
In the process of genetic inheritance, humans develop gametes, which consist of only one chromosome of each of the 23 pairs and thus are haploid. The two gametes – one male and one female – combine to a zygote which is again diploid and builds the origin of the offspring. The meiosis, i.e., the development of gametes, is a two-step procedure and begins with a diploid cell containing two copies of each chromosome. During the first phase of the meiosis the pairs of homologous chromosomes separate into two cells. In the second phase each chromosome is decoupled into two chromatids which segregate into daughter cells forming four haploid gametes. Before the chromosomes are passed on to the daughter cells, crossovers may occur between two homologous chromosomes at random positions. An uneven number of crossovers within an interval between two gene loci on the same chromosome of a gamete leads to a recombination between these two loci.[9]

1.2. Genomewide Linkage Studies

Genomewide linkage analysis is a family based method to identify chromosomal regions which harbor genetic variants responsible for inherited diseases. For this type of disease mapping, highly polymorphic genetic markers, usually microsatellites, which are frequent and widespread across the whole genome, are used for investigating co-segregation with the disease in a sample of families with parents and offspring or in a pedigree of many generations. Linkage analysis can also be carried out with SNVs, however, their lower degree of information has to be compensated by an increased number of variants.
Linkage analysis compares the observed inheritance pattern of two loci, e.g., the trait locus and a genetic marker, within a pedigree. An independent inheritance speaks in favor of unlinked loci or free recombination.[9] Then the probability of recombination between the two loci, also called recombination fraction, is $\theta = 0.5$. A co-segregation indicates linkage between those two loci, $\theta < 0.5$. Thus in a linkage analysis of two loci the hypothesis $H_0 : \ \theta = 0.5$ is tested against $H_1 : \ \theta < 0.5$.[9] This analysis is usually performed with a putative trait locus and all typed markers on a genomewide scale.
Linkage analyses have been very successful in the identification of genes responsible for Mendelian disorders;[10] however mapping genes causing complex traits is more difficult

since they often are influenced by additional genetic and environmental factors.[11] Moreover, epistasis, incomplete penetrance, epigenetic effects and phenotypic heterogeneity hamper the identification of genes causing complex diseases.

1.3. Association Studies

Mendelian disorders are typically caused by rare mutations, however most of the individual genetic variation is due to common variants.[12] One of the reasons is natural selection against mutations that are strongly deleterious, which holds for most Mendelian diseases. For common diseases however, a theory exists that most likely common variants, defined as having a minor allele frequency >1%, are responsible for common diseases[13,14] – the so-called "common disease–common variant" (CD-CV) hypothesis. Common disorders often exhibit late onset with moderate or no obvious impact on reproductive fitness.[12] Therefore, mildly deleterious alleles can gain moderate frequencies, particularly in populations that have undergone recent expansion.[15] Alleles having a beneficial or neutral effect during the process of human evolution might now confer disease susceptibility because of environmental changes. Finally, disease causing alleles can maintain a high frequency if they are additionally responsible for a beneficial phenotype. An example is the sickle-cell mutation in the beta hemoglobin gene (*HBB*) which confers malaria resistance to heterozygous carriers.[16]
In recent years SNPs have become the main source of studying genetic variation in humans.[17] The methods of choice in family studies are transmission disequilibrium based tests (TDT)[18] whereas in studies of unrelated individuals, including case-control studies, frequency- and regression-based methods have been applied.
With the advent of GWAS, the preferred study design has switched to case-control studies. The comparison of marker allele frequencies between affected and unaffected individuals is the basis of genetic association studies. In an initial map of 1.42 million SNPs, the average SNP density was estimated as one per 1.9 kilobases.[19] More recently, approximately one SNP per 300 base pairs in the human genome was expected.[12] Meanwhile the map has grown to 38 million SNPs and together with short insertions and deletions it covers about 98% of the genetic variation at a frequency of at least 1%.[20]
Risch and Merikangas[14] argue that association studies – as opposed to linkage studies – provide a more powerful tool for detecting common disease risk variants, even when contributing only modestly to the disease risk. Until the start of the era of GWAS in 2005, association studies focused on candidate genes, which were selected through putative involvement in disease pathogenesis due to their function or their location in regions identified by genomewide linkage studies.[21] However, candidate studies often turned out to be unsuccessful. The limitation of studying certain variants within biologically plausible genes was that each had a small prior probability of being causal, and the chance of false positive results due to inhomogeneous population structure, which could not be adequately addressed at that time, was high. The latter was one of the major pitfalls,[12] but it only applies to the case-control designs since the TDT is robust against population stratification as the non-transmitted parental alleles serve as internal controls.[18] Moreover,

clinical heterogeneity of the investigated trait can limit statistical power. Inappropriate selection of control groups and the use of inadequate statistical models led to only few consistently replicated findings in the candidate gene approach.[22,23,24]
For atopic eczema (AE), regarding candidate-gene studies only the association with the filaggrin gene (*FLG*) first described by Palmer et al.[25] could be consistently replicated in more than 20 studies conducted after the first replication was carried out by our working group.[26]

1.3.1. Genomewide Association Studies (GWAS)

With denser publicly available maps, more insight into the genetic architecture has been gained. Variants in close proximity often form a block-like structure, building a region of little or no recombination. Thus these regions of SNPs in strong linkage disequilibrium (LD) with each other limit the haplotype diversity[27,28] and allow researchers to infer untyped SNPs. The haplotype patterns were built by recombinational hot and cold spots[29,30] as well as historical population bottlenecks.[31] Hot spots refer to locations where crossover events cluster, i.e., those loci exhibit higher recombination rates than on average in the genome, and cold spots show recombination rates below average. Commonly used statistical measurements for pairwise LD are D' and r^2 with values in the range between 0 and 1. Both measures are based on the linkage disequilibrium coefficient D, which can be derived from Table 1.1.
Let p and q be the allele frequencies of SNP_A and SNP_B and $f_{A_1B_1}$,...,$f_{A_2B_2}$ the observed gametic frequencies, then $D = f_{A_1B_1} - pq$, the difference between observed and expected gametic frequencies. Standardizations lead to the LD-measures $r^2 = \frac{D^2}{p(1-p)q(1-q)}$ and $D' = \frac{D}{D_{max}}$, where

$$D_{max} = \begin{cases} min(p(1-q);\,(1-p)q) & \text{if } D > 0 \\ max(-pq;\,-(1-p)(1-q)) & \text{if } D < 0 \end{cases}$$

r^2 is a genetic analogon to the square of the Pearson correlation coefficient whereas D' assesses the probability of historical recombination.[32] In contrast to D', the r^2 measure depends additionally on the allele frequencies at both loci.
This block-like organization of the genome has some implications on the analysis strategies. Hence, a limited set of LD-based tagging SNPs might suffice for disease mapping. Given the block-like structure, as little as 500,000 or 1,000,000 SNPs need to be genotyped in European populations, and twice this number for African populations, to cover most of the genetic variation.[28,33] Moreover, imputation techniques allow for inferring non-genotyped SNPs by using an adequate reference population and MCMC-based estimation techniques as implemented among others in MACH,[34] IMPUTE[35] and SHAPE-IT.[36]
Assuming the CD-CV hypothesis, genomewide association studies (GWAS) of common variants were expected to be most useful for mapping genes responsible for common diseases to specific loci.
Public data from the International HapMap Project[37] and the 1000 Genomes Project,[38] launched in 2002 and 2010, respectively, and the development of chip-based high-throughput

		Allele at SNP_B		
		B_1	B_2	
Allele at SNP_A	A_1	$f_{A_1B_1}$ pq	$f_{A_1B_2}$ $p(1-q)$	p
	A_2	$f_{A_2B_1}$ $(1-p)q$	$f_{A_2B_2}$ $(1-p)(1-q)$	$1-p$
		q	$1-q$	1

Table 1.1.: Association between two SNPs. Observed gametic frequencies ($f_\bullet$) and expected gametic frequencies (product of the allele frequencies when linkage equilibrium exists).

genotyping at affordable costs have made it possible to analyze up to 4 million SNPs simultaneously (Illumina Infinium HumanOmni5-Quad BeadChip[39]).
Genomewide association studies have been successful in recent years. They have identified numerous new and often replicated susceptibility loci for many complex diseases,[40] e.g. type 2 diabetes, Crohn's disease, or asthma, and represent a very useful tool for further dissection of complex traits. However, GWAS depend on an appropriate study design in terms of sample size, adequate selection of cases and controls, control for population stratification, stringent statistical analysis and reproducibility of results. Moreover, results of association between certain SNPs and traits should be interpreted with caution, since associations most likely appear indirectly through LD with the functional variant that has often not been genotyped on genomewide chips.
In the field of atopic diseases, our working group identified susceptibility loci (*FCER1A* on chromosome 1 and *RAD50* on chromosome 5) for total IgE[41] and a locus in an intergenic region between *C11orf30* and *LRRC32* which is putatively involved in the epithelial differentiation associated with AE.[42]

1.3.2. Meta Analysis

The idea of summarizing results from different studies on gambling dates back to the 18th and 19th century where mathematicians like Gauss and Laplace drew attention to errors and combination of observations. However, it was not until the 20th century that statisticians worked on similar questions as to summarize results from different clinical trials.[43]
In 1904 Pearson[44] combined studies on typhoid fever and compared vaccinated vs. non-vaccinated soldiers in different places across the British Empire. He grouped them with respect to similarity of their (geographical) origin. For each of his sub-studies he presented the correlation of inoculation with infection and mortality, respectively, along with their probable errors as a measure of within-study uncertainty. Additionally he calculated a

mean correlation value as a summary statistic across all sub-studies, which can be seen as a form of 'meta-analysis'.
Ronald A. Fisher had a strong impact on meta-analysis and he "encouraged scientists to summarize their research in such a way to make the comparison and combination of estimates almost automatic, and the same as if all the data were available".[43]
Cochran,[45] who extended Fisher's approach, assumed unequal standard errors across study centers and provided a formal framework for a random effects model. He showed that for less than 15 study centers the weighted mean is not an efficient estimate for the mean response and should be replaced by a maximum likelihood estimate.
However, only by the large amount of research reports in the middle of the 20th century scientists started to think about methods of synthesizing results.[43]
First, psychologists drew attention to this problem, and in 1976 Glass[46] referred the term *meta-analysis* to "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings."
In medical research meta-analyses were applied a few years later, mainly influenced by randomized trials on aspirin vs. placebo, which found only a suggestive but not statistically significant mortality reduction in patients treated with aspirin after a heart attack. However, a meta-analysis of the six conducted trials confirmed the suggested findings.[43]
Peto and his colleagues stimulated researchers with their work to perform systematic reviews of clinical trials and to combine estimates of treatment effects based on "direct, unweighted summation of O(bserved)-E(xpected) value from each trial".[47]
Meta-analysis has become an integral part of systematic reviews with a clear set of rules for study search, inclusion and exclusion criteria in many research fields.[48]
The Cochrane collaboration is a worldwide non-profit, non-governmental network of "health practitioners, scientists, patient advocates and others" and has become a driving force in systematically organizing medical research information in order to "promote evidence-informed health decision-making".[49] It has been recognized as "a long-standing, rigorous, and innovative leader in developing methods" for meta-analysis.[50] Based on the principles of evidence-based medicine, the Cochrane collaboration conducts publicly-available, high-quality, relevant and up-to date research synthesis. Furthermore, it develops and provides guidelines for systematic reviews[51] including protocols of structured literature search as well as "new and extended analytic and diagnostic methods for evaluating the output of meta-analysis".[50] Recently, the PRISMA statement has been published which consists of a checklist and a flow diagram to enhance the reporting of meta-analysis and systematic reviews.[52]

Fixed Effects Model

For combining K studies in a meta-analysis, from each study $k = 1, \ldots, K$ the estimate of the effect size of interest $\hat{\beta}_k$ and its standard error $\hat{\sigma}_k$ has to be retrieved. The effect sizes can be e.g. correlations, which will be first transformed to Fisher's z-scale,[53] or based on means for continuous outcomes. Risk difference, log-transformed risk ratio or odds ratio are commonly chosen for binary responses and log-transformed hazard ratio for time-to-event data.

The fixed effects model assumes that there is a common true effect size θ shared by all included studies and the variation between studies is a result of the random sampling error. Thus the overall estimate $\hat{\theta}$ is calculated by

$$\hat{\theta} = \frac{\sum w_k \hat{\beta}_k}{\sum w_k} \tag{1.1}$$

where w_k are the weights for each study calculated by the inverse variance method $w_k = \frac{1}{\hat{\sigma}_k^2}$ and the sum is taken over studies $k = 1, \ldots, K$. The overall variance is calculated by

$$Var(\hat{\theta}) = \frac{1}{\sum w_k} \tag{1.2}$$

Under the assumption that the overall effect size is normally distributed,[54] confidence intervals can be constructed by

$$\mathrm{CI}_{\hat{\theta}} = \hat{\theta} \pm 1.96 \cdot \sqrt{Var(\hat{\theta})}. \tag{1.3}$$

Finally, a Z-score can be computed to test the null hypothesis whether the common true effect θ is zero

$$\begin{aligned} Z &= \frac{\hat{\theta}}{\sqrt{Var(\hat{\theta})}} \\ P &= (1 - \Phi(|Z|)) \cdot 2 \end{aligned} \tag{1.4}$$

with $\Phi(\cdot)$ being the cumulative distribution function of the standard normal distribution.

Random Effects Model

Like Cochran in the 1930s,[45] Armitage[55] stressed on thorough consideration of how to "draw inference from heterogenous but logically related studies". DerSimonian and Laird[56] picked up the idea of heterogeneity between studies and provided formulae for estimating the true variation in effect sizes τ^2 using the methods of moment (MOM), maximum likelihood (ML) or restricted maximum likelihood estimates (REML). Thus the variance of the effect estimate for each study $\hat{\sigma}_k^{2*}$ is divided in two components, the between-study variance and the within-study variance i.e., $\hat{\sigma}_k^{2*} = \hat{\tau}^2 + \hat{\sigma}_k^2$. Then the overall estimate and variance are calculated by

$$\begin{aligned} \hat{\theta}_{DL} &= \frac{\sum w_k^* \hat{\beta}_k}{\sum w_k^*} \\ \mathrm{Var}(\hat{\theta}_{DL}) &= \frac{1}{\sum w_k^*} \end{aligned} \tag{1.5}$$

with weights $w_k^* = 1/\hat{\sigma}_k^{2*}$. DerSimonian and Laird showed that the ML estimate was biased downwards and while there was little difference between MOM and REML, MOM is the method of choice since it does not require iterations and is computed as

$$\hat{\tau}^2_{DL} = \max\left(0, \frac{Q-(K-1)}{\sum w_k - \frac{\sum w_k^2}{\sum w_k}}\right) \tag{1.6}$$

with

$$Q = \sum w_k(\hat{\beta}_k - \hat{\theta})^2 \tag{1.7}$$

Until today the MOM estimate by of DerSimonian and Laird[56] has been widely applied and has been often referred to as the DL-method. Since then, several modifications[57,58,59] or more sophisticated and iterative methods[60,61,62] for estimating τ^2 have been proposed. Several simulation studies have shown that the DL-method is appropriate in most scenarios and computer-intensive alternatives only slightly improve the estimate of the overall effect θ.[63,64,65,66] However, when the focus is on significance testing regarding θ, alternative methods have been shown to perform better than the DL-method, since there is no clear null hypothesis.[67] The DL-method has been criticized as being too conservative if under H_0 τ^2 is zero[68] and too liberal if under H_0 τ^2 is > 0.[69] Since both components, the within $\hat{\sigma}^2_k$ and between $\hat{\tau}^2$ study variance, are estimated, and the calculated weights w^*_k are assumed to be fixed and known, the DL-method disregards the variability of the weights and may provide too small variance estimates which leads to a too liberal test.[67] However, the DL-method mostly has an approximately correct type 1 error rate as long as the number of studies and the within-study samples are large.[67] Hartung & Knapp[58] and Sidik & Jonkman[59] independently proposed a refined estimate for $\text{Var}(\theta)$ based on weighted least squares:

$$\text{Var}(\hat{\theta}_{HS}) = \frac{\sum w^*_k(\hat{\beta}_k - \hat{\theta})^2}{(K-1)\sum w^*_k} \tag{1.8}$$

Both argue that the resulting test statistic $\hat{\theta}/\sqrt{\text{Var}(\hat{\theta}_{HS})}$ is t-distributed with $K-1$ degrees of freedom. The refinement has been shown to provide a more accurate test.[69,70] Finally, an ad hoc modification for the variance estimate is recommended in case of few studies contributing to the meta-analysis or varying standard errors $\hat{\sigma}^*_k$ by[71]

$$\text{Var}(\hat{\theta}) = \max\left(\frac{1}{\sum w^*_k}, \text{Var}(\hat{\theta}_{HS})\right).$$

In order to decide whether the fixed or random effects model should be applied, researchers often test the homogeneity assumption of effect sizes based on the Q-statistic (1.7), which follows a χ^2_{K-1}-distribution under H_0. As all significance tests, this test is sensitive to the magnitude and precision of the effects, and conclusions about heterogeneity should not solely be drawn based on Q,[72] because it has low power in case of few studies and does not report the magnitude of heterogeneity. Therefore the I^2 statistic has been proposed, which reflects the amount of variance on a relative scale

$$I^2 = \left(\frac{Q-df}{Q}\right)\cdot 100 \tag{1.9}$$

and can be interpreted as a ratio of between-study variance to the total variance.[72] If there is no between-study variance τ^2, the random and fixed effects model are the same, and apart from special cases, the DL method, which takes random effects into account, remains a valid and powerful approach.

Meta-Regression

Meta-regression is a valuable generalization of meta-analysis that allows for incorporating the effect of one or several study characteristics and can be helpful in exploring sources of heterogeneity.[67] To this end classical (multiple) regression methods can be applied. In contrast to primary study analysis, covariates are defined on study level rather than individual level, and the response is the study effect size instead of the subject's outcome. Further a weight to each study has to be assigned and researchers have to choose between the random or fixed effects model. In analogy to primary studies, for an appropriate meta-regression a reasonable ratio of study to covariates is needed, and in case of few studies meta-regression is not recommended.[73] Significance tests on the slope parameter(s) are based on the normality assumption under H_0, and for assessing the impact of several covariates simultaneously an analogon to the F-test, the Q-test, is defined.[73]
In analogy to meta-analysis, the distinction of fixed and random effects in meta-regression is also based on the assumption of the effect size. While the fixed effects model assumes one population effect size, the random effects model assumes "a distribution of effect sizes for studies with the same predicted value (e.g. studies share the same values on all covariates)".[73] In order to estimate the between-study variance τ^2 in meta-regression, extensions of the MOM (1.6) and REML method are available. Knapp and Hartung[74] proposed a refined variance estimator in analogy to (1.8) adapted for the regression model and showed that their test with the refined variance estimator outperforms the commonly used test in terms of holding the type 1 error rate.
In practice, some issues have to be considered for meta-regression, such as an appropriate choice of covariates, which should be pre-specified and restricted to a certain number also taking their functional relation with the study effects into account.[75] Finally, meta-regression can serve as a useful technique to explore hidden heterogeneity, but the results have to be interpreted with caution.

Further Extensions

In case of a categorical covariate the meta-regression can also be performed with methods "akin to analysis of variance in a primary study"[76] by using models for fixed and random effects.
Meta-analysis can be further extended to summarize studies of complex data structure, e.g. multiple independent subgroups (e.g. assess a treatment on mild, moderate or severe AE), or multiple outcomes or time-points within a study.
For analysis of multiple independent subgroups three options have been proposed:[77]

- Treat each subgroup (e.g. mild, moderate or severe AE) of each study as an independent study and calculate the overall effect by a meta-analysis.

- Compute a combined effect for each study by meta-analyzing across subgroups within each study. Then meta-analyze the summary statistics of all studies.
- Compute a combined effect for each subgroup across all studies by meta-analysis (e.g. summary statistics for mild, moderate and severe AE) and then combine the resulting summary statistics by a further meta-analysis.

At each meta-analysis step researchers can choose between fixed or random effects models, whereas the latter usually is more appropriate.
For combination of multiple outcomes or time-points it has been suggested to compute a common effect and its variance across outcomes per study, incorporating the correlation between the outcomes. In each study, the common effect can either be a summary effect of multiple outcomes or the difference for contrasting two outcomes. Subsequently a meta-analysis across studies will be carried out by using a fixed or random effects model.[78] Similar analysis strategies can be applied to combine studies with e.g. several treatment groups within a study.
In the field of randomized trials an extension of systematic reviews, the network meta-analysis, became popular, which addresses the question how to rank competing treatment regimens.[79] To this end, treatments are organized in a network as nodes with edges defining pairwise treatment comparisons, and observed relations are reflected in the design matrix. Two parametrizations of the network are proposed and the simultaneous synthesis exploits direct and indirect evidence of particular treatment comparisons. While the contrast-based parametrization models the difference between trial arms, the arm-based parametrization describes the mean outcome of each treatment. For both parametrizations a fixed and a random effects model can be employed. Although the arm-based parametrization has been criticized as being not identifiable, it has recently been shown that it delivers comparable results to the contrast based parametrization and yields some advantages such as simplifying random effects modelling, allowing treatment-specific random effect variance, and direct estimation of reference treatment outcome.[80]
Finally, extensions have been developed to relax the consistency assumption, i.e., the transitivity among treatment effects between direct and indirect evidence.[81] For example, recently, a simple, fast and direct method to fit a network meta-analysis with a random inconsistency effect was proposed based on the DL-method.[82]
In conclusion, multivariate meta-analysis may be beneficial in certain situations and "can provide estimates with better statistical properties", however it should be carried out with more caution than its univariate counterpart.[83]

Publication Bias

Sterling first thought about publication bias and noted in his work[84] that because non-significant results are scarcely published and other researchers are not aware of these investigations, it is likely that false positive results, significant by chance, preferentially find their way into the literature. On the other hand, published results meeting the significance threshold of $P < 0.05$ will hardly be challenged by an independent investigation.[84]
An early method, the *Rosenthal's Fail-safe N*, calculates how many "unobserved studies"

would be needed with effect size zero until the overall combined P-value is above the significance threshold.[85] An extension of this approach, the *Orwin's Fail-safe N*, addresses two criticisms of the early approach, which is based on statistical rather than clinical relevance and assumes that the missing studies have an effect size of zero.[86] However, these methods are not recommended for Cochrane reviews.[87]
Funnel plots are a graphical way of assessing a potential publication bias by looking at the symmetry in a scatter plot of effect size vs. a measure of study size, e.g. number of individuals, standard error σ_k^2 or the transformation $1/\sigma_k^2$. The underlying assumption is that effect sizes should be distributed symmetrically at any fixed level of σ_k^2. Deviation from symmetry might indicate a publication bias, since large studies are likely to be published irrespective of their significance, moderately sized studies might be lost if not significant, and small studies are at greater risk of being lost and will only be published with significant large effect sizes. Funnel plots should be assessed with caution, since visual inspection is largely subjective and may lead to wrong conclusions. Egger et al.[88] introduced a simple test of the asymmetry of this plot by regressing the standardized treatment effect (z-score: $\frac{\hat{\beta}_k}{\hat{\sigma}_k}$) on the inverse squared standard error $1/\sigma_k^2$. However, this test should be applied with caution, because it only makes sense with a reasonable number of studies and variation in study size, and even then the test is likely to be underpowered.[86] An alternative method, which is also based on the symmetry of the funnel plot, is "trim and fill", which iteratively removes the extreme studies and estimates an overall effect $\hat{\theta}_{\text{trunk}}$ of the symmetric part of the plot after truncation. Then it adds the removed studies back to the analysis with their unobserved counterparts. To this end, the unobserved studies are imputed by using $\hat{\theta}_{\text{trunk}}$ and the observed estimates $\hat{\beta}_k$ of the asymmetric part.[89] The full augmented data are used to estimate the overall θ. The "trim and fill" method should be used for sensitivity analysis rather than finding the number of missing studies.[90] However there is much debate whether tests based on symmetry are sufficient to conclude a publication bias, since the tests tend to be underpowered and even when asymmetry is observed publication bias is not necessarily present.[91] In conclusion, methods for detecting publication bias should be applied with caution and results should not be over-interpreted in any way.

Meta-Analysis of Genomewide Association Studies

Recently, the statistical concept of meta-analysis entered the field of genomewide association studies. Zeggini et al.[92] and the DIAGRAM Consortium conducted one of the first meta-analyses on type II diabetes where they combined three GWAS comprising over 4,500 cases and over 5,500 controls. In contrast to a classical meta-analysis based on literature search, in GWAS researchers form consortia and contribute the results of their individual study. The studies have to be harmonized since individual studies are usually not genotyped on the same platform and hence different SNP sets have been investigated. Harmonization is carried out by means of imputation on the same reference panel, using the same strand annotation and with similar quality control procedures sorting out badly genotyped and imputed SNPs before resulting test statistics are combined. Imputation stands for estimating non-genotyped SNPs based on the haplotype structure of the cor-

responding reference panel. Moreover, meta-analysis in the context of GWAS aims at hypothesis testing rather than estimation of effect parameters,[93] since the causal variant remains a priori unknown and the main goal is mapping complex diseases to genes. Since pooling the large amount of data from all studies is often not feasible in practice, meta-analysis is a convenient tool to increase power for detecting small genetic effects, "but only if the degree of heterogeneity is small."[67]
Meta-analyses of GWAS, especially in cohorts of European descent, strongly contributed to our understanding of complex trait genetics. An increasing number of non-European GWAS drew attention on how to synthesize data across diverse populations in order to increase power for detecting novel loci.[94] Recently, a method for transethnic meta-analysis was proposed, which allows clustering of populations with similar allelic effects while taking account of heterogeneity between groups of diverse ancestry.[95] This method can be seen as a compromise between a fixed-effects meta-analysis assuming the same allelic effect in all populations and the random-effects meta-analysis assuming a different allelic effect in each population.
Our working group conducted two meta analyses on the effect of mutations in the *FLG* gene on AE.[96,97] Additionally we have been participating in the large EArly Genetics and Lifecourse Epidemiology (EAGLE) Eczema consortium, which summarized in 2011 several European-descent GWAS studies on AE with over 5,000 affected individuals and 20,000 controls and identified two novel loci associated with AE.[98] Recently, in a multi-ancestry meta-analysis of 26 genome-wide association studies of over 21,000 cases and 95,000 controls imputed to the 1000 Genomes Project Phase 1 reference panel (release March 2012) the EAGLE consortium identified another ten new risk loci. This collaborative effort increased the total number of of known AE risk loci to 31, which explain 14.9% of the estimated heritability.[99]

1.4. Objectives

Genetic epidemiology research has lately focused on fine-mapping established GWAS loci. To this end, customized arrays have been developed. For example the Immunochip was designed by a consortium to better define risk variants of immune-mediated diseases. The analysis revealed pleiotropic loci also associated with AE.[3] Antagonistic and agonistic effects of these loci were observed regarding AE, e.g. with autoimmune diseases such as asthma, allergic rhinitis, Crohn's disease, and psoriasis. The interest in genetic overlap, which helps to identify common pathways, is steadily increasing.
In this work the statistical concept of meta-analysis is introduced and a new meta-analysis based approach, the **C**ompare & **C**ontrast **M**eta-**A**nalysis (CCMA), is proposed, which allows researchers to categorize effects into disease-specific, agonistic or antagonistic. In order to establish the method as a means to inferentially test for pleiotropy, the distribution of the test statistic under the null hypothesis (H_0) is explored and an approximate analytical distribution is determined. The power of the proposed method is evaluated by simulation studies and compared with the Subset-Based Meta-Analysis (SBMA) method.[7]
As an application, psoriasis and AE are compared and contrasted by exploiting three

GWAS datasets for each disease. The identified significant SNVs are compared with the results obtained by a multinomial regression model as gold standard as well as by SBMA and the COMPARED & OVERLAP approach.[6] Finally, the pleiotropic loci identified for psoriasis and AE are shown in detail, the results are interpreted in the light of current knowledge of the genetic architecture of both diseases, and conclusions are made by discussing the proposed method in a general context.

1.5. Contributing Manuscripts

Parts of this thesis are published in peer-reviewed journals, which have been written in cooperation with my supervisor Konstantin Strauch, the head of my working group Stephan Weidinger, as well as with colleagues and collaborators in the field of statistics and genetic epidemiology.

The method section 2.3 and the corresponding results section 3.1 have been published in

> Baurecht H, Hotze M, Rodríguez E, Manz J, Weidinger S, Cordell HJ, Augustin T, Strauch K (2016): Compare and Contrast Meta Analysis (CCMA): A Method for Identification of Pleiotropic Loci in Genome-Wide Association Studies. *PLoS One* 5;11(5):e0154872.

All analytical work and simulation studies were carried out by Hansjörg Baurecht. The manuscript was written by Hansjörg Baurecht under the supervision of Konstantin Strauch. All other authors critically reviewed the manuscript.

The results sections 3.3 - 3.6 have been published in

> Baurecht H, Hotze M, Brand S, Büning C, Cormican P, Corvin A, Ellinghaus D, Ellinghaus E, Esparza-Gordillo J, Fölster-Holst R, Franke A, Gieger C, Hubner N, Illig T, Irvine AD, Kabesch M, Lee YA, Lieb W, Marenholz I, McLean WH, Morris DW, Mrowietz U, Nair R, Nöthen MM, Novak N, O'Regan GM; Psoriasis Association Genetics Extension., Schreiber S, Smith C, Strauch K, Stuart PE, Trembath R, Tsoi LC, Weichenthal M, Barker J, Elder JT, Weidinger S, Cordell HJ, Brown SJ (2015): Genome-wide comparative analysis of atopic dermatitis and psoriasis gives insight into opposing genetic mechanisms. Am J Hum Genet. 2015 Jan 8;96(1):104-20.

The test statistic was developed by Hansjörg Baurecht and simulation studies were conducted with the help from Melanie Hotze. The statistical analysis of comparing and contrasting atopic eczema and psoriasis was carried out by Hansjörg Baurecht and Melanie Hotze. The manuscript was written by Hansjörg Baurecht and Sarah J Brown under the supervision of Heather J Cordell, Stephan Weidinger and Konstantin Strauch. All other authors contributed data and critically reviewed the manuscript.

2. Methods

The **C**ompare & **C**ontrast **M**eta-**A**nalysis (CCMA) is a modification of the idea of Bhattacharjee et al.,[7] who proposed a meta-analysis based method for combining GWAS results of heterogeneous traits to identify common pleiotropic effects. To this end, the method of Bhattacharjee et al. explores all subsets of studies for the strongest association signals, which may be in the same or opposite direction, followed by an evaluation of the significance of the signal accounting for multiple testing. In contrast to standard meta analysis it allows some subset of studies to exhibit no effects.

2.1. Meta Analysis Revisited

As already described, meta analysis is a method to synthesize results from different studies. Therefore from each study $k = 1, \ldots, K$ the estimates $\hat{\beta}_k$ and corresponding standard errors $\hat{\sigma}_k$ are combined to an overall estimate $\hat{\theta}$ and variance $\mathrm{Var}(\hat{\theta})$ by using a fixed or random effects model.
For fixed effects models it is assumed that the effects of each study represents a random realization of the *true* overall effect. As outlined in subsection 1.3.2, the overall estimate $\hat{\theta}$ and its variance are calculated by formulae (1.1) and (1.2).
As previously mentioned, the random effects model relaxes the fixed effects assumption and heterogeneity between studies is allowed. It is assumed that each study has its own *true* effect and these effects are normally distributed around an overall effect. The heterogeneity variance between studies τ^2 is usually estimated by the DL-method (1.6).[56] The variance of each study is given by $\hat{\sigma}_k^{2*} = \hat{\tau}^2 + \hat{\sigma}_k^2$ and thus their weights by $w_k^* = \frac{1}{\hat{\sigma}_k^{2*}}$. Then the overall effect estimate and its variance are calculated by formulae (1.5).

In the field of genetics, the fixed-effects meta-analysis assumes the same allelic effect in all cohorts, while the random-effects model assumes a different underlying allelic effect for each population. In populations of the same ancestry, e.g. cohorts of European descent, the fixed-effect assumption is reasonable. It has been argued that the random-effects model is too conservative, and a hybrid method as a compromise for transethnic meta-analysis has been proposed.[95]
In homogeneous populations an alternative approach to calculate genetic association is justifiable and similar to the fixed-effects model. Instead of combining the studies' effect estimates and their standard errors by the inverse variance method as described in

subsection 1.3.2, each study's P-value P_k will be transformed into a signed Z-score by[100]

$$Z_k = \Phi^{-1}\left(\frac{P_k}{2}\right) \cdot \text{sign}(\Delta_k) \tag{2.1}$$

where $\Phi^{-1}(\cdot)$ denotes the inverse of the cumulative distribution function of the standard normal distribution and Δ_k the effect direction of study k. Thus, "very negative Z-scores indicate a small P-value and an allele associated with lower disease risk or quantitative trait levels, whereas large positive Z-scores indicate a small P-value and an allele associated with higher disease risk or quantitative trait levels".[100] Then the overall test statistic is calculated by

$$Z_{meta} = \frac{\sum Z_k w_k}{\sqrt{\sum w_k^2}} \tag{2.2}$$

where the weights w_k can be defined in different ways:[7,100]

$$\begin{aligned} w_k &= \sqrt{N_k} \\ w_k &= \frac{1}{\hat{\sigma}_k} \Big/ \frac{1}{\sqrt{\sum 1/\hat{\sigma}_k^2}} \\ w_k &= \sqrt{\pi_k} \end{aligned} \tag{2.3}$$

N_k denotes the sample size of study k and $\pi_k = N_k / \sum N_k$ is the relative study size.

2.2. Subset Based Meta Analysis

The Subset Based Meta Analysis (SBMA) exploits GWAS results using the estimates $\hat{\beta}_k$ and their standard errors $\hat{\sigma}_k$ and allows some studies to have no effect.[7] For any given subset S of the studies the Z-statistic is calculated as

$$Z(S) = \sum_{k \in S} \sqrt{\pi_k(S)} Z_k \tag{2.4}$$

where $\pi_k(S) = N_k / \sum_{k \in S} N_k$ is the relative sample size of study k to the total sample size of subset S.
The overall significance for an association of an SNV with the disease is calculated by

$$Z_{max-meta} = T = \max_{S \in \mathbf{S}} |Z(S)| \tag{2.5}$$

i.e., the maximum Z-statistic of subset S over all possible subsets $\mathbf{S}$ with $|\mathbf{S}| = 2^K - 1$ is chosen. The authors use the discrete local maxima method[101] (DLM), which approximates tail probabilities of a multivariate normally distributed Z-statistic maximized over a non-smooth grid and thus gives an improved Bonferroni-type bound for the P-value. To this end, the authors construct a grid by defining subsets $S \pm k$ as neighbors. The neighbors differ in the k^{th} study, which is added or removed, depending on whether the k^{th} study is included in the current subset S or not. The DLM exploits the fact that the event $\{|Z(S)| > T\}$ is contained in the union of the events

$\{|Z(S)| > T\}$ and {for all neighboring subsets $Z(S \pm k) < Z(S)$}. The approximate one-sided P-value can be calculated as follows:

$$\tilde{P}_{\text{DLM}} = \sum_{S \in \mathbf{S}} \int_T^\infty 2 \cdot \left\{ \prod_k^K Pr(|Z(S \pm k)| < z | Z(S) = z) \right\} \phi(z) dz \tag{2.6}$$

where $\phi(.)$ denotes the standard normal density and $S \pm k$ denotes the neighbor of the current subset S obtained by including or excluding the k^{th} study. Under the assumption of a separable correlation structure between the neighboring subsets while conditioning on the test statistic $Z(S)$ for subset S, Markov property is invoked, which implies that the test statistics for all neighboring subsets $S \pm k$ are conditionally independent given the statistic for the current subset $Z(S)$. The P-value is calculated by evaluating the conditional probability separately for each neighboring set $Z(S \pm k)$ using univariate conditional standard normal distribution. To this end, the authors argue that conditional on $Z(S) = z$, $|Z(S+k)| < z$ is equivalent to $Z(k) < u_k(z)$ and $|Z(S-k)| < z$ is equivalent to $Z(k) > l_k(z)$. The terms $u_k(z)$ and $l_k(z)$ denote the upper $\frac{(1-w_S)z}{w_k}$ and lower bound $\frac{(w_S-1)z}{w_k}$, respectively, and w_S and w_k the respective study weights, which are defined as follows:

$$w_k = \sqrt{\pi_k(S+k)} = \sqrt{\frac{N_k}{\sum_{s \in (S+k)} N_s}}$$
$$w_S = \sqrt{\pi_S(S+k)} = \sqrt{\frac{\sum_{s \in S} N_s}{\sum_{s \in (S+k)} N_s}}$$

Thus $\tilde{P}_{\text{DLM}}$ can be written as

$$\begin{aligned} \tilde{P}_{\text{DLM}} &= \sum_{S \in \mathbf{S}} \int_T^\infty 2 \cdot \left\{ \prod_k^K Pr(l_k(z) < Z(k) < u_k(z) | Z(S) = z) \right\} \phi(z) dz \\ &= \sum_{S \in \mathbf{S}} \int_T^\infty 2 \cdot \left\{ \prod_k^K \Omega_k \right\} \phi(z) dz \end{aligned} \tag{2.7}$$

with weights $\Omega_k = Pr(l_k(z) < Z(k) < u_k(z) | Z(S) = z)$. The product of weights $\prod_k^K \Omega_k$ is used for down-weighting the tails of the standard normal distribution and yields a corrected P-value contribution for each subset. The Ω_k can be evaluated with a univariate conditional normal distribution.[7] Therefore, the covariances between a pair of subsets is defined by a function of the number of cases and controls. It allows researchers to take into account the potential overlap of individuals between the two subsets. An example for the case of two studies is shown in Supplement A.1.
Basically, the $\tilde{P}_{\text{DLM}}$ sums up the corrected tail probabilities of all subsets. To this end, for each subset a product of weights determined by the neighboring subsets $S \pm k$ is calculated

under the assumption that they are conditionally independent given the current subset S. The resulting factor down-weights the tail probability of the current subset in order to improve the Bonferroni bound

$$P = \sum_{S \in \mathbf{S}} \int_T^\infty \phi(z) dz$$

by incorporating more information. Since under the null hypothesis each statistic $Z(S)$ for any subset $S \in \mathbf{S}$ is asymptotically standard normally distributed, the evaluation of the Ω_k as well as the integration of the weighted standard normal density function in eq. (2.7) is justified.
For detecting effects in opposing directions, the authors propose a two-sided test by separately calculating the one-sided P-value above for subsets of studies showing association signals in positive and negative directions and thus obtain $Z_{\max+}$ and $Z_{\max-}$, respectively. This is equivalent to

$$\begin{aligned} Z_{\max+} &= \max_{S \in \mathbf{S}^+} Z(S) \\ Z_{\max-} &= \max_{S \in \mathbf{S}^-} Z(S) \end{aligned} \tag{2.8}$$

where $\mathbf{S}^+$ and $\mathbf{S}^-$ denote all studies showing a positive or negative test-statistic, respectively. Then the P-values for $\tilde{P}^+_{\text{DLM}}$ and $\tilde{P}^-_{\text{DLM}}$ are combined using Fisher's method

$$Z_{\text{max-meta}^{(2)}} = -2 \left\{ \log \tilde{P}^+_{\text{DLM}} + \log \tilde{P}^-_{\text{DLM}} \right\} \tag{2.9}$$

$$\tilde{P}^{(2)}_{\text{DLM}} = P\left(\chi^2_4 > Z^{(2)}_{\text{max-meta}} \right). \tag{2.10}$$

In Appendix A.1 $\tilde{P}_{\text{DLM}}$ is explicitly calculated for the case of two independent studies, which corresponds to the problem of comparing and contrasting two different traits.

2.3. Compare & Contrast Meta Analysis (CCMA)

In the CCMA method the GWAS results of each disease are combined by a classical meta-analysis, since there is no search for an optimal subset of studies in an agnostic way. The method intends to compare and contrast both diseases in order to find pleiotropic loci that show effects in the same or opposite direction. Thus, the studies for each disease are meta-analyzed separately assuming a fixed effects model and $\hat{\theta}$ and $\text{Var}(\hat{\theta})$ are calculated by using Equations (1.1), (1.2). An overall test statistic for the first (T_1) and second disease (T_2) is derived using

$$T_i = \frac{\hat{\theta}_i}{\sqrt{\text{Var}(\hat{\theta}_i)}} \tag{2.11}$$

where $\hat{\theta}_i$ is the overall effect for disease $i = 1, 2$ and $\text{Var}(\hat{\theta}_i)$ its estimated variance. Then the CCMA test statistic T_{max} is calculated by[102]

$$T_{max} = \max(|T_1|, |T_2|, |T_{12,\text{agonistic}}|, |T_{12,\text{antagonistic}}|) \tag{2.12}$$

where $T_{12,\text{agonistic}}$ and $T_{12,\text{antagonistic}}$ are defined as

$$T_{12,\text{agonistic}} = \frac{T_1 + T_2}{\sqrt{2}} = \frac{\sqrt{\text{Var}(\hat{\theta}_2)}\hat{\theta}_1 + \sqrt{\text{Var}(\hat{\theta}_1)}\hat{\theta}_2}{\sqrt{\text{Var}(\hat{\theta}_1)}\sqrt{\text{Var}(\hat{\theta}_2)}\sqrt{2}}$$

$$T_{12,\text{antagonistic}} = \frac{T_1 - T_2}{\sqrt{2}}.$$

Under the homogeneity (fixed effects) assumption the meta-analysis test statistics T_1 and T_2 asymptotically follow a standard normal distribution under H_0.[103] By design, $T_{12,\text{agonistic}}$ and $T_{12,\text{antagonistic}}$ also asymptotically follow a standard normal distribution, utilizing the properties of the sum and difference of independent normally distributed random variables. In order to derive a P-value for an observed realization t_{max}, the null distribution is empirically determined by simulating two independent normally distributed random variables $Z_1 \sim N(0,1)$ and $Z_2 \sim N(0,1)$ of R =1,000,000,000 replicates. Then

$$Z_{12,\text{agonistic}} = \frac{Z_1 + Z_2}{\sqrt{2}}$$

and

$$Z_{12,\text{antagonistic}} = \frac{Z_1 - Z_2}{\sqrt{2}}$$

as well as the test statistic for the empirical null distribution

$$Z_{max} = \max(|Z_1|, |Z_2|, |Z_{12,\text{agonistic}}|, |Z_{12,\text{antagonistic}}|) \tag{2.13}$$

are calculated for each replicate. The empirical P-values can be derived as

$$P_{emp} = \frac{\#(Z_{max} > t_{max}) + 1}{R + 1} \tag{2.14}$$

In order to find an analytic formulation of the P-value distribution, the squared values of the test statistics $Z_1^2, Z_2^2, Z_{12,\text{agonistic}}^2, Z_{12,\text{antagonistic}}^2$ under H_0, corresponding to no pleiotropy and no association between the SNV and any trait, are considered, for which the CCMA test statistic is constructed. By design, each of the four transformed variables follows a χ_1^2 distribution with $Z_1^2 \perp Z_2^2$ and $Z_{12,\text{agonistic}}^2 \perp Z_{12,\text{antagonistic}}^2$ under H_0 (see Section 2.3.2). Thus, the transformed CCMA test statistic can be expressed by

$$Z_{max}^2 = \max(Z_1^2, Z_2^2, Z_{12,\text{agonistic}}^2, Z_{12,\text{antagonistic}}^2) \tag{2.15}$$

and empirical P-values can be calculated for an observed realization by

$$P_{emp} = \frac{\#(Z_{max}^2 > t_{max}^2) + 1}{R + 1} \tag{2.16}$$

Plotting $-log_{10}(P_{emp})$ against Z_{max}^2 suggests that the relationship can be expressed by a straight line (Figure 2.1).

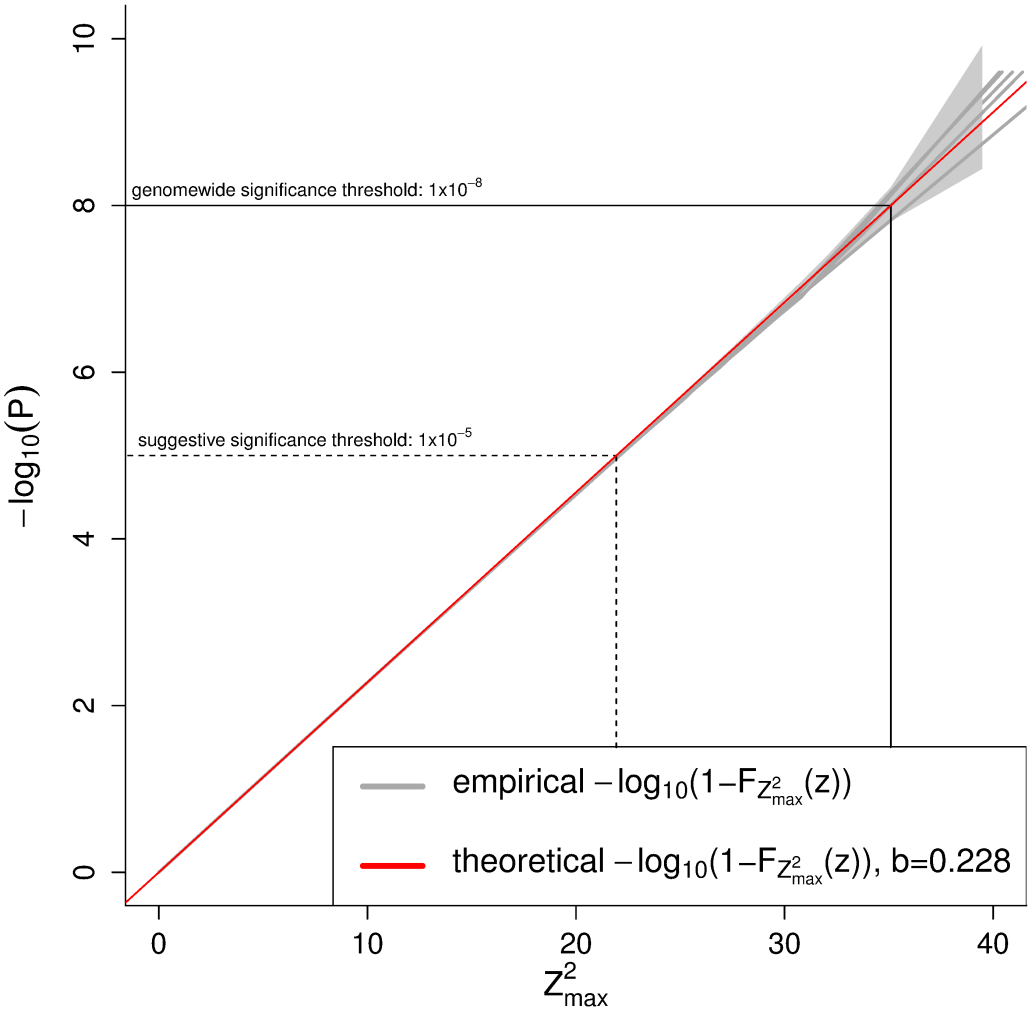


Figure 2.1.: Five empirical evaluations of $-log_{10}(P)$-distribution of the $Z^2_{\max}$ statistic, each obtained by simulating 2×10^9 replicates. The theoretical distribution was obtained by fitting a straight line with the estimated slope parameter $b = 0.228$. The grey shaded area reflects the 95% Clopper-Pearson confidence interval.[104]

A general formula for the distribution and density function of the maximum of independent identically-distributed (iid) variables has been described in chapter 2.11 by Ewens & Grant.[105] Let $X_1, X_2, \ldots, X_k$ be continuous iid variables and $X_{\max} = \max(X_1, X_2, \ldots, X_k)$ their maximum, then the cumulative distribution function of $X_{\max}$ can be written as follows:

$$\begin{aligned} P(X_{\max} \leq x) &= P(X_1 \leq x \cap X_2 \leq x \cap \cdots \cap X_k \leq x) = \{P(X \leq x)\}^k \\ &= F_{X_{\max}}(x) = \{F_X(x)\}^k \end{aligned} \tag{2.17}$$

In the case of the CCMA test statistic formula (2.17) cannot be applied directly, since there are no four independent variables. However, these variables can be divided into two blocks each with iid χ_1^2-distributed variables $Z_1^2 \perp Z_2^2$ and $Z_{12,\text{agonistic}}^2 \perp Z_{12,\text{antagonistic}}^2$. Let $F_{\chi_1^2}(z)$ denote the distribution function of each variable $Z_1^2, Z_2^2, Z_{12,\text{agonistic}}^2, Z_{12,\text{antagonistic}}^2$ and $F_{Z_{\max}^{2*}}(z)$ denote the distribution function of $Z_{\max}^{2*} = \max(Z_1^2, Z_2^2)$ or $Z_{\max}^{2*} = \max(Z_{12,\text{agonistic}}^2, Z_{12,\text{antagonistic}}^2)$, then

$$F_{Z_{\max}^{2*}}(z) = \left\{ F_{\chi_1^2}(z) \right\}^2 \tag{2.18}$$

Furthermore it is known that the sum of two iid χ_1^2-distributed variables is χ_2^2-distributed with the cumulative distribution function $F_{\chi_2^2}(z)$. Since for the CCMA test statistic there are only two independent random variables Z_1^2 and Z_2^2, the following boundaries for $F_{Z_{\max}^2}(z)$ may be postulated:

$$F_{Z_{\max}^{2*}}(z) \geq F_{Z_{\max}^2}(z) \geq F_{\chi_2^2}(z) \tag{2.19}$$

To prove that $F(x) \geq F(y)$, it has to be shown that $x \leq y$. It can be seen that $\max(Z_1^2, Z_2^2) \leq Z_1^2 + Z_2^2$ and thus $F_{Z_{\max}^{2*}}(z) \geq F_{\chi_2^2}(z)$. Further, it is obvious that $\max(Z_1^2, Z_2^2) \leq \max\left(Z_1^2, Z_2^2, \frac{(Z_1+Z_2)^2}{2}, \frac{(Z_1-Z_2)^2}{2}\right)$ and therefore $F_{Z_{\max}^{2*}}(z) \geq F_{Z_{\max}^2}(z)$. Finally, $F_{Z_{\max}^2}(z) \geq F_{\chi_2^2}(z)$ is proven by showing that $\max\left(Z_1^2, Z_2^2, \frac{(Z_1+Z_2)^2}{2}, \frac{(Z_1-Z_2)^2}{2}\right) \leq Z_1^2 + Z_2^2$. Since obviously $Z_1^2 \leq Z_1^2 + Z_2^2$ and $Z_2^2 \leq Z_1^2 + Z_2^2$ it remains to be shown that $\frac{(Z_1+Z_2)^2}{2} \leq Z_1^2 + Z_2^2$ and $\frac{(Z_1-Z_2)^2}{2} \leq Z_1^2 + Z_2^2$.

$$\begin{aligned}
\frac{(Z_1 + Z_2)^2}{2} &\leq Z_1^2 + Z_2^2 \\
\frac{Z_1^2 + 2Z_1Z_2 + Z_2^2}{2} &\leq Z_1^2 + Z_2^2 \\
\frac{Z_1^2 + Z_2^2}{2} + Z_1Z_2 &\leq Z_1^2 + Z_2^2 \\
Z_1Z_2 &\leq \frac{(Z_1^2 + Z_2^2)}{2} \\
Z_1^2 - 2\,Z_1Z_2 + Z_2^2 &\geq 0 \\
(Z_1 - Z_2)^2 &\geq 0 \quad \text{(agonistic case)}
\end{aligned} \tag{2.20}$$

This obviously holds irrespective of the values of the test statistics Z_1^2 and Z_2^2. For the antagonistic case, only the sign in the left term has to be changed, which results in

$$(Z_1 + Z_2)^2 \geq 0 \quad \text{(antagonistic case)} \tag{2.21}$$

which also holds for any values of Z_1^2 and Z_2^2. This concludes the proof of equation (2.19). Therefore, with formula (2.19) explicit boundaries for $F_{Z_{\max}^2}(z)$ are established, which are visualized in Figure 2.2.

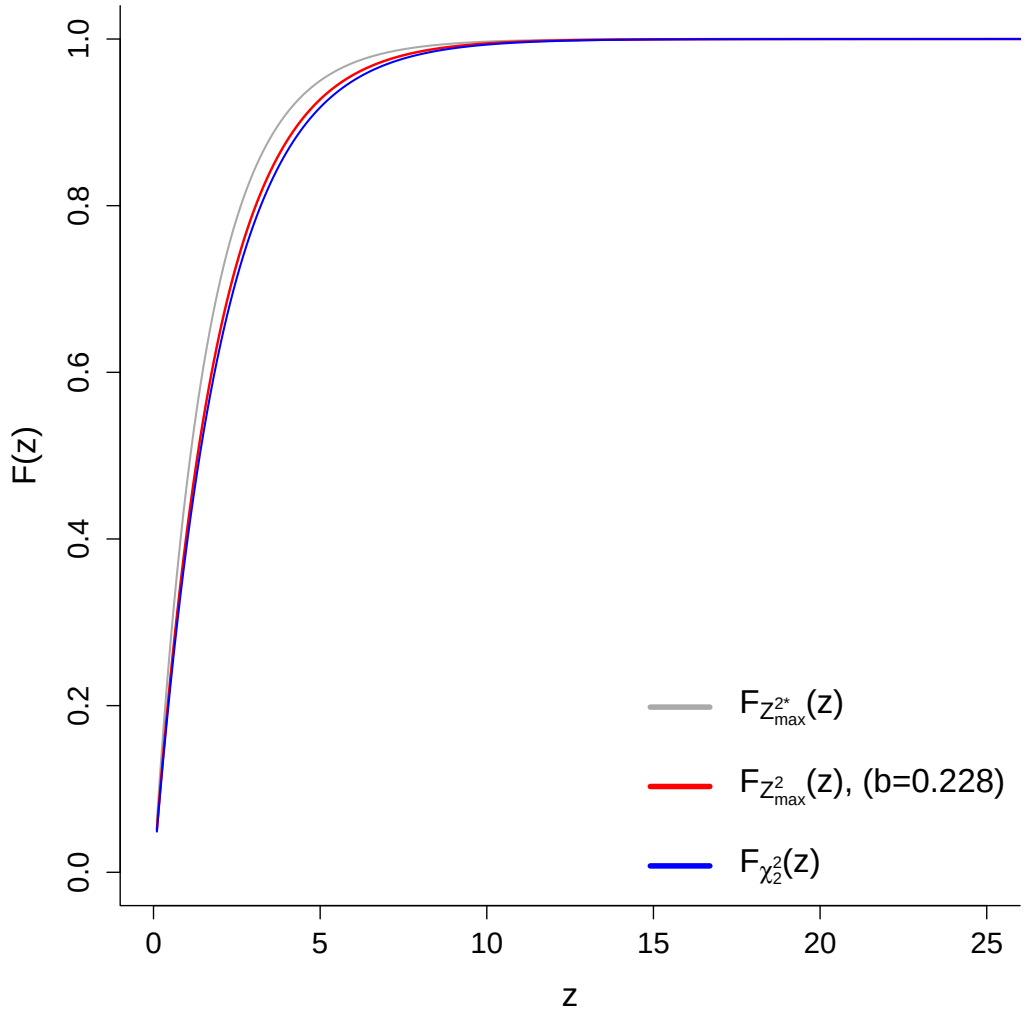


Figure 2.2.: Comparison of theoretical distributions $F_{Z^{2*}_{\max}}(z)$, $F_{Z^2_{\max}}(z)$ (with fitted parameter $b = 0.228$) and $F_{\chi^2_2}(z)$.

It turns out to be important that $F_{Z^2_{\max}}(z)$ is exponentially distributed. To deduce this, note that $F_{\chi^2_2}(z)$ can be expressed in terms of an exponential distribution $F_\lambda(z)$ with scale parameter $\lambda = \frac{1}{2}$

$$F_\lambda(z) = 1 - e^{-\lambda \cdot z} \tag{2.22}$$

and $F_\lambda(z)$ can be connected to z by a linear relation

$$F_\lambda(z) = 1 - e^{-\lambda \cdot z} \iff -log(1 - F_\lambda(z)) = \lambda \cdot z \tag{2.23}$$

Given that the relationship of $-log_{10}(P)$ and $Z^2_{\max}$ is a straight line under H_0 (Figure 2.1), the cumulative distribution function of $Z^2_{\max}$ is

$$\begin{aligned} -log_{10}(P) &= b \cdot z \\ P &= 10^{-b \cdot z} \\ F_{Z^2_{\max}}(z) = (1 - P) &= 1 - 10^{-b \cdot z} \end{aligned} \tag{2.24}$$

Using the relationship $10^x = e^{log(10) \cdot x}$, $F_{Z^2_{\max}}(z)$ can be written as an exponential distribution

$$\begin{aligned} F_{Z^2_{\max}}(z) &= 1 - 10^{-b \cdot z} \\ &= 1 - e^{-log(10) \cdot b \cdot z} \\ &= 1 - e^{-\lambda z} \text{ with } \lambda = log(10) \cdot b \end{aligned} \tag{2.25}$$

In order to determine the theoretical distribution, the optimal slope parameter b has to be found. To this end, two simulations have been conducted, one of 100 empirical $Z^2_{\max}$ distributions with R=1,000,000,000 and one of 5 empirical $Z^2_{\max}$ with R=2,000,000,000 replicates. The slope parameter is estimated by means of linear regression and a consistent estimate of $b \approx 0.228$ is found (Table 2.1).

Setting	Min	Q1	Median	Q3	Max	Mean	Std Dev
100 sim.with 1×10^9 repl.	0.22786	0.22795	0.22797	0.2280	0.22809	0.22797	$3.88 \cdot 10^{-5}$
5 sim. with 2×10^9 repl.	0.22796	0.22797	0.22798	0.22798	0.22799	0.22798	$1.08 \cdot 10^{-5}$

Table 2.1.: Distribution of the slope parameter b of simulated $Z^2_{\max}$ distributions by different simulation settings. sim.=simulations, repl.=replicates, Q1=25% quantile, Q3=75% quantile.

With (2.24) and (2.25) we can give a formula for the cumulative distribution function of the original (not squared) $Z_{\max}$ statistic:

$$\begin{aligned} F_{Z_{\max}}(z) &= 1 - 10^{-b \cdot z^2} \\ &= 1 - e^{-log(10) \cdot b \cdot z^2} \end{aligned} \tag{2.26}$$

Finally, formula (2.26) represents the cumulative distribution function of the original $Z_{\max}$ statistic and we compare it with its simulated values from the previous study.[106] Theoretical thresholds for suggestive (10^{-5}) and genomewide (10^{-8}) significance of $Z_{\max} = 4.68$ and $Z_{\max} = 5.92$ are found (Figure 2.3). These thresholds correspond well to the values of 4.7 and 6 derived by our previous simulation studies (see Methods section in Baurecht et al.[106]).

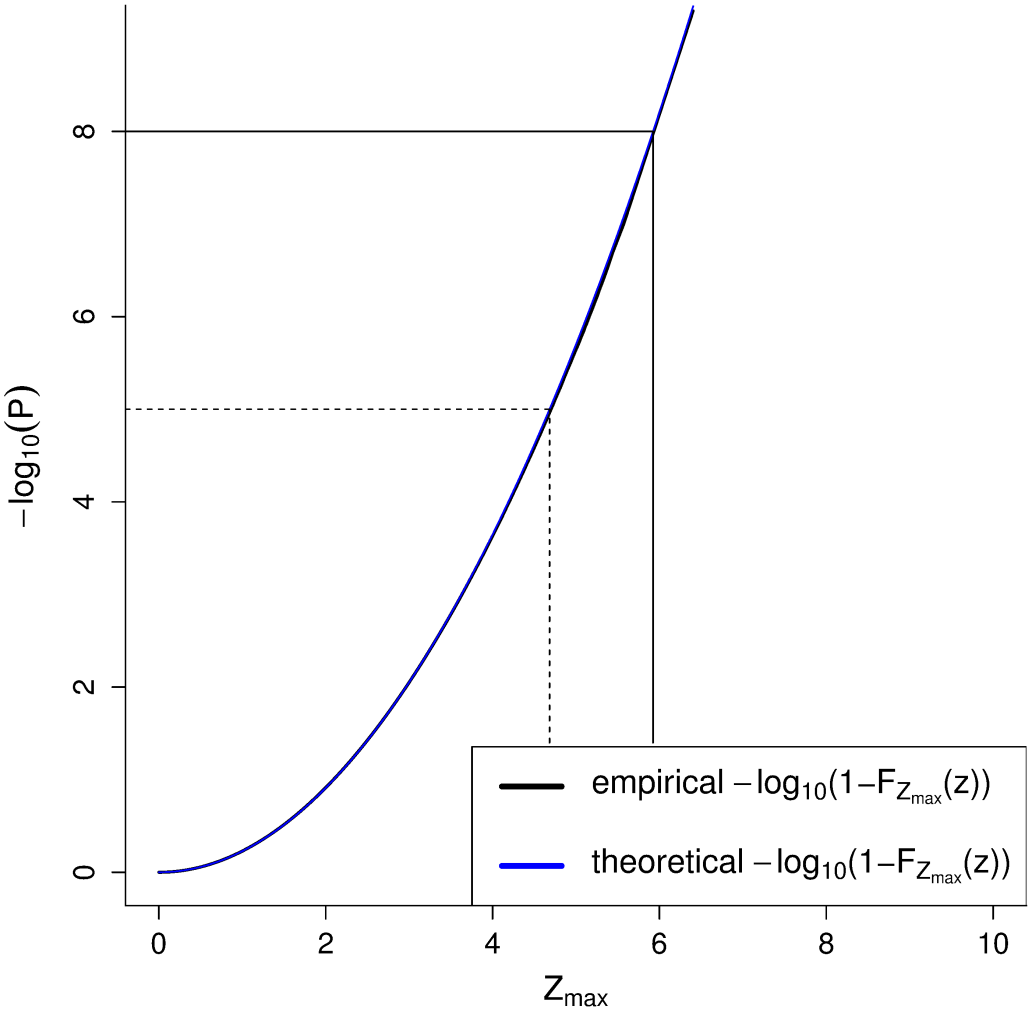


Figure 2.3.: Empirical and theoretical $-log_{10}(P)$-distribution of $Z_{\max}$ with fitted parameter $b = 0.228$. Dotted and solid grey lines indicate the thresholds of suggestive ($Z_{\max} = 4.68$) and genomewide significance ($Z_{\max} = 5.92$).

2.3.1. Weighted CCMA Test Statistic (wCCMA)

A minor modification of the CCMA test statistic, termed wCCMA, allows to incorporate weights taking into account the study size.

Let $w_k = \sqrt{\frac{N_k}{N}}$ be the weights for study k and $\mathbf{x} = (Z_1, Z_2)^T$. Then the transformation matrix $\mathbf{A}$ can be specified as $\mathbf{A}_{(1)} = \frac{1}{\sqrt{w_1^2+w_2^2}} \begin{pmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{pmatrix}$. Again, it can be shown that the transformed statistics $\mathbf{y} = (Z_{12,\text{agonistic}}, Z_{12,\text{antagonistic}})^T = \mathbf{A}_{(1)}\mathbf{x}$ are independent (Section 2.3.2).

Alternatively, the transformation matrix $\mathbf{A}$ can be specified as $\mathbf{A}_{(2)} = \frac{1}{\sqrt{w_1^2+w_2^2}} \begin{pmatrix} w_2 & w_1 \\ w_1 & -w_2 \end{pmatrix}$. Subsequently, both modified versions are termed wCCMA^1 and wCCMA^2.

2.3.2. Proof of Independence between $Z_{12,\text{agonistic}}$ and $Z_{12,\text{antagonistic}}$

To prove that $Z_{12,\text{agonistic}} \perp Z_{12,\text{antagonistic}}$, we show that $\mathbf{y} = (Z_{12,\text{agonistic}}, Z_{12,\text{antagonistic}})^T \sim N_2(\mu_\mathbf{y}, \mathbf{\Sigma_y})$ with the identity matrix $\mathbf{\Sigma_y} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Using the theorem (3.4) in Fahrmeir et al.[107] on page 26

Theorem 1 *Let* $\mathbf{x} \sim N_p(\mu, \mathbf{\Sigma})$ *and* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$*, in which* $\mathbf{A}$ *is a* $(q \times p)$*-matrix with* $rk(\mathbf{A}) = q \leq p$*. Then* $\mathbf{y} \sim N_q(\mathbf{A}\mu, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T)$*.*

we let $\mathbf{x} = (Z_1, Z_2)^T$ be $N_2(\mu_\mathbf{x}, \mathbf{\Sigma_x})$ distributed with $\mu_\mathbf{x} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ and $\mathbf{\Sigma_x} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then, $\mathbf{y} = (Z_{12,\text{agonistic}}, Z_{12,\text{antagonistic}})^T = \mathbf{A}\mathbf{x} + \mathbf{b}$ with $\mathbf{A} = \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$ and $\mathbf{b} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$.
It can be seen that $E(\mathbf{y}) = \mu_y = \mathbf{A}\mu_\mathbf{x} + \mathbf{b} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ with variance

$$
\begin{aligned}
\mathbf{\Sigma_y} &= \mathbf{A}\mathbf{\Sigma_x}\mathbf{A}^T \\
&= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \frac{1}{\sqrt{2}} \\
&= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\end{aligned}
$$

For the modified version of the CCMA test statistic the independence of the transformed statistics $\mathbf{y} = (Z_{12,\text{agonistic}}, Z_{12,\text{antagonistic}})^T = \mathbf{A}_{(1)}\mathbf{x} + \mathbf{b}$ can also be shown. For wCCMA[1] the transformation matrix is defined as $\mathbf{A}_{(1)} = \frac{1}{\sqrt{w_1^2+w_2^2}}\left(\begin{smallmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{smallmatrix}\right)$, $\mathbf{b} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$, and again it can bee seen that $E(\mathbf{y}) = \mu_\mathbf{y} = \mathbf{A}_{(1)}\mu_\mathbf{x} + \mathbf{b} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ with variance

$$
\begin{aligned}
\mathbf{\Sigma_y} &= \mathbf{A}_{(1)}\mathbf{\Sigma_x}\mathbf{A}_{(1)}^T \\
&= \frac{1}{\sqrt{w_1^2 + w_2^2}} \begin{pmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{pmatrix} \frac{1}{\sqrt{w_1^2 + w_2^2}} \\
&= \frac{1}{w_1^2 + w_2^2} \begin{pmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{pmatrix} \begin{pmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\end{aligned}
$$

For wCCMA[2] the transformation matrix is specified as $\mathbf{A}_{(2)} = \frac{1}{\sqrt{w_1^2+w_2^2}}\left(\begin{smallmatrix} w_2 & w_1 \\ w_1 & -w_2 \end{smallmatrix}\right)$, $\mathbf{b} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$, and as above it can be seen that $E(\mathbf{y}) = \mu_\mathbf{y} = \mathbf{A}_{(2)}\mu_\mathbf{x} + \mathbf{b} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ with variance

$$
\begin{aligned}
\mathbf{\Sigma_y} &= \mathbf{A}_{(2)}\mathbf{\Sigma_x}\mathbf{A}_{(2)}^T \\
&= \frac{1}{\sqrt{w_1^2 + w_2^2}} \begin{pmatrix} w_2 & w_1 \\ w_1 & -w_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w_2 & w_1 \\ w_1 & -w_2 \end{pmatrix} \frac{1}{\sqrt{w_1^2 + w_2^2}} \\
&= \frac{1}{w_1^2 + w_2^2} \begin{pmatrix} w_2 & w_1 \\ w_1 & -w_2 \end{pmatrix} \begin{pmatrix} w_2 & w_1 \\ w_1 & -w_2 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\end{aligned}
$$

2.3.3. Construction of Confidence Intervals

Let $(X_1, \ldots, X_n)$ be a sample of independent and identically distributed random variables with $X_i \sim F_\theta$. Then confidence intervals can be constructed by using the following rules presented on page 107 in Spodarev:[108]

1. Find a test statistic $T(X_1, \ldots, X_n; \theta)$ which depends on the parameter θ and has a known distribution F under $\mathbb{P}_\theta$, which itself does not depend on θ (see Kabluchko[109] page 84).
2. Define quantiles $F^{-1}(\alpha_1)$ and $F^{-1}(1-\alpha_2)$ of the distribution F for the levels α_1 and $1-\alpha_2$, such that the overall level $\alpha = \alpha_1 + \alpha_2$.
3. Solve the inequation $F^{-1}(\alpha_1) \leq T(X_1, \ldots, X_n; \theta) \leq F^{-1}(1-\alpha_2)$ with respect to θ.

By doing so, a confidence interval $I = \left[T_\theta^{-1}(F^{-1}(\alpha_1)),\ T_\theta^{-1}(F^{-1}(1-\alpha_2))\right]$ for the true parameter value θ for the level of $1-\alpha$ can be obtained if T_θ is a monotone increasing function, since

$$\begin{aligned}
\mathbb{P}_\theta(\theta \in I) &= \mathbb{P}_\theta\left(T_\theta^{-1}(F^{-1}(\alpha_1)) \leq \theta \leq T_\theta^{-1}(F^{-1}(1-\alpha_2))\right) \\
&= \mathbb{P}_\theta\left(F^{-1}(\alpha_1) \leq T_\theta(X_1, \ldots, X_n, \theta) \leq F^{-1}(1-\alpha_2)\right) \\
&= F\left(F^{-1}(1-\alpha_2)\right) - F\left(F^{-1}(\alpha_1)\right) \\
&= 1 - \alpha_2 - \alpha_1 \\
&= 1 - \alpha \qquad \forall\, \theta \in \Theta
\end{aligned} \tag{2.27}$$

(see Spodarev[108] page 107).
As shown earlier in this section, $Z^2_{\max}$ follows an exponential distribution under H_0 (eq. 2.25) with $\lambda_0 = log(10) \cdot b$ with $b \approx 0.228$. $\theta_{\max}$ can be defined as the maximum effect deviating from the null hypothesis. Then, the CCMA test statistic $T^2_{\max}$ (eq. 2.12) can be written depending on $\theta_{\max}$ as

$$\begin{aligned}
T^2_{\max,\theta_{\max}} &= \left(\max\left(|T_1|, |T_2|, |T_{12,\text{agonistic}}, |T_{12,\text{antagonistic}}|\right) - \theta_{\max}\right)^2 \\
&= \left(T_{\max} - \theta_{\max}\right)^2
\end{aligned} \tag{2.28}$$

which has a known distribution $F_{Z^2_{\max}} = Exp(\lambda)$ with $\lambda = log(10) \cdot 0.228$.
Confidence intervals for the maximum effect $\theta_{\max}$ can be constructed by applying the rules from above and using the distribution function given in (2.25) with the estimated distribution parameter $b = 0.228$. By design the test statistic $T^2_{\max,\theta_{\max}}$ has positive values, thus a $1-\alpha$ confidence interval can be constructed with

$$\begin{aligned}
1 - \alpha &= \mathbb{P}_{\theta_{\max}}\left(T^2_{\max,\theta_{\max}} \leq F^{-1}_{Z^2_{\max}}(1-\alpha)\right) \\
&= \mathbb{P}_{\theta_{\max}}\left(\left(T_{\max} - \theta_{\max}\right)^2 \leq F^{-1}_{Z^2_{\max}}(1-\alpha)\right) \\
&= \mathbb{P}_{\theta_{\max}}\left(-\sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)} \leq T_{\max} - \theta_{\max} \leq \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)}\right)
\end{aligned}$$

The inequalities $-\sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)} \leq T_{\max} - \theta_{\max}$ and $T_{\max} - \theta_{\max} \leq \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)}$ will be solved with respect to $\theta_{\max}$

$$\begin{aligned} T_{\max} - \theta_{\max} &\leq \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)} \\ \theta_{\max} &\geq T_{\max} - \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)} \end{aligned}$$

$$\begin{aligned} T_{\max} - \theta_{\max} &\geq -\sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)} \\ \theta_{\max} &\leq T_{\max} + \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)} \end{aligned}$$

Then the $1-\alpha$ confidence interval for the unknown maximum effect $\theta_{\max}$ is given as

$$I = \left[T_{\max} - \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)},\ T_{\max} + \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)}\right] \tag{2.29}$$

As defined in eq. (2.12), the CCMA test statistic is the maximum of the absolute values of the four statistics T_1, T_2, $T_{12,\text{agonistic}}$ and $T_{12,\text{antagonistic}}$:

$$T_1 = \frac{\hat{\theta}_1}{\sqrt{\text{Var}(\hat{\theta}_1)}}$$

$$T_2 = \frac{\hat{\theta}_2}{\sqrt{\text{Var}(\hat{\theta}_2)}}$$

$$T_{12,\text{agonistic}} = \frac{T_1 + T_2}{\sqrt{2}} = \frac{\sqrt{\text{Var}(\hat{\theta}_2)}\hat{\theta}_1 + \sqrt{\text{Var}(\hat{\theta}_1)}\hat{\theta}_2}{\sqrt{2\,\text{Var}(\hat{\theta}_1)\text{Var}(\hat{\theta}_2)}}$$

$$T_{12,\text{antagonistic}} = \frac{T_1 - T_2}{\sqrt{2}} = \frac{\sqrt{\text{Var}(\hat{\theta}_2)}\hat{\theta}_1 - \sqrt{\text{Var}(\hat{\theta}_1)}\hat{\theta}_2}{\sqrt{2\,\text{Var}(\hat{\theta}_1)\text{Var}(\hat{\theta}_2)}} \tag{2.30}$$

with $\hat{\theta}_1$ and $\hat{\theta}_2$ being the effect estimates obtained from a meta-analysis for disease 1 and disease 2 (see section 2.3). $\sqrt{\text{Var}(\hat{\theta}_1)}$ and $\sqrt{\text{Var}(\hat{\theta}_2)}$ are the standard errors of the respective effect estimates.
Finally, for the effect of interest (disease specific, agonistic or antagonistic), depending on which of the four statistics gives the maximum, a confidence interval can be specified

by inserting the four test statistics in (2.29) and solving for the respective numerator in (2.30):

$$I_1 = \left[\hat{\theta}_1 - \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)\,\mathrm{Var}(\hat{\theta}_1)},\ \hat{\theta}_1 + \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)\,\mathrm{Var}(\hat{\theta}_1)}\right]$$

$$I_2 = \left[\hat{\theta}_2 - \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)\,\mathrm{Var}(\hat{\theta}_2)},\ \hat{\theta}_2 + \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)\,\mathrm{Var}(\hat{\theta}_2)}\right]$$

$$\begin{aligned} I_{12,\text{agonistic}} = \Bigg[&\left(\sqrt{\mathrm{Var}(\hat{\theta}_2)}\hat{\theta}_1 + \sqrt{\mathrm{Var}(\hat{\theta}_1)}\hat{\theta}_2\right) - \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)\,2\,\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)},\\ &\left(\sqrt{\mathrm{Var}(\hat{\theta}_2)}\hat{\theta}_1 + \sqrt{\mathrm{Var}(\hat{\theta}_1)}\hat{\theta}_2\right) + \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)\,2\,\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}\Bigg] \end{aligned}$$

$$\begin{aligned} I_{12,\text{antagonistic}} = \Bigg[&\left(\sqrt{\mathrm{Var}(\hat{\theta}_2)}\hat{\theta}_1 - \sqrt{\mathrm{Var}(\hat{\theta}_1)}\hat{\theta}_2\right) - \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)\,2\,\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)},\\ &\left(\sqrt{\mathrm{Var}(\hat{\theta}_2)}\hat{\theta}_1 - \sqrt{\mathrm{Var}(\hat{\theta}_1)}\hat{\theta}_2\right) + \sqrt{F^{-1}_{Z^2_{\max}}(1-\alpha)\,2\,\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}\Bigg] \end{aligned} \tag{2.31}$$

The transformed confidence interval in eq. (2.31) for disease specific effects is wider than a classical confidence interval using the estimated effect $\hat{\theta}$, its standard error and the normal approximation.[54] It takes into account the greater variability by searching for the maximum test statistic. For agonistic or antagonistic effects the resulting confidence intervals are difficult to interpret, since the respective numerator in eq. (2.30) is a weighted sum or difference of the disease specific effects.
Deviding the numerator and the denominator of $T_{12,\text{agonistic}}$ and $T_{12,\text{antagonistic}}$ in eq. (2.30) by $\sqrt[4]{\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}$ converges the weights in the respective numerator towards one as can be seen in eq. (2.32). In the case of equal variances the numerator reflects exactly the sum or difference of $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively.

$$T_{12,\text{agonistic}} = \frac{\frac{\sqrt{\mathrm{Var}(\hat{\theta}_2)}\hat{\theta}_1 + \sqrt{\mathrm{Var}(\hat{\theta}_1)}\hat{\theta}_2}{\sqrt[4]{\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}}}{\frac{\sqrt{2\,\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}}{\sqrt[4]{\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}}} = \frac{\sqrt[4]{\frac{\mathrm{Var}(\hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1)}}\hat{\theta}_1 + \sqrt[4]{\frac{\mathrm{Var}(\hat{\theta}_1)}{\mathrm{Var}(\hat{\theta}_2)}}\hat{\theta}_2}{\sqrt{2}\sqrt[4]{\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}}$$

$$T_{12,\text{antagonistic}} = \frac{\frac{\sqrt{\mathrm{Var}(\hat{\theta}_2)}\hat{\theta}_1 - \sqrt{\mathrm{Var}(\hat{\theta}_1)}\hat{\theta}_2}{\sqrt[4]{\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}}}{\frac{\sqrt{2\,\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}}{\sqrt[4]{\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}}} = \frac{\sqrt[4]{\frac{\mathrm{Var}(\hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1)}}\hat{\theta}_1 - \sqrt[4]{\frac{\mathrm{Var}(\hat{\theta}_1)}{\mathrm{Var}(\hat{\theta}_2)}}\hat{\theta}_2}{\sqrt{2}\sqrt[4]{\mathrm{Var}(\hat{\theta}_1)\mathrm{Var}(\hat{\theta}_2)}} \tag{2.32}$$

This leads to an approximate confidence interval for the agonistic or antagonistic effects, which is depicted in eq. (2.33):

$$I_{12,\text{agonistic}} = \left[\left(\sqrt[4]{\frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}}\hat{\theta}_1 + \sqrt[4]{\frac{\text{Var}(\hat{\theta}_1)}{\text{Var}(\hat{\theta}_2)}}\hat{\theta}_2\right) - \sqrt{2\,F^{-1}_{Z^2_{\max}}(1-\alpha)}\,\sqrt[4]{\text{Var}(\hat{\theta}_1)\text{Var}(\hat{\theta}_2)},\right.$$
$$\left.\left(\sqrt[4]{\frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}}\hat{\theta}_1 + \sqrt[4]{\frac{\text{Var}(\hat{\theta}_1)}{\text{Var}(\hat{\theta}_2)}}\hat{\theta}_2\right) + \sqrt{2\,F^{-1}_{Z^2_{\max}}(1-\alpha)}\,\sqrt[4]{\text{Var}(\hat{\theta}_1)\text{Var}(\hat{\theta}_2)}\right]$$

$$I_{12,\text{antagonistic}} = \left[\left(\sqrt[4]{\frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}}\hat{\theta}_1 - \sqrt[4]{\frac{\text{Var}(\hat{\theta}_1)}{\text{Var}(\hat{\theta}_2)}}\hat{\theta}_2\right) - \sqrt{2\,F^{-1}_{Z^2_{\max}}(1-\alpha)}\,\sqrt[4]{\text{Var}(\hat{\theta}_1)\text{Var}(\hat{\theta}_2)},\right.$$
$$\left.\left(\sqrt[4]{\frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}}\hat{\theta}_1 - \sqrt[4]{\frac{\text{Var}(\hat{\theta}_1)}{\text{Var}(\hat{\theta}_2)}}\hat{\theta}_2\right) + \sqrt{2\,F^{-1}_{Z^2_{\max}}(1-\alpha)}\,\sqrt[4]{\text{Var}(\hat{\theta}_1)\text{Var}(\hat{\theta}_2)}\right] \tag{2.33}$$

In general, it is possible to define for each confidence interval a corresponding statistical test, but not vice versa (see Spodarev[108] page 107).

2.4. Power and Type I Error Analysis

The power and type I error of the CCMA method is compared with those of the SBMA approach (two-sided test $\tilde{P}^{(2)}_{\text{DLM}}$) implemented in the R-package ASSET[110,111] by simulations. To this end, a fixed population of n=20,000 individuals is generated with the respective genotype distribution according to the specified minor allele frequency (MAF) in exact Hardy-Weinberg equilibrium (HWE). Then, n=8,000 individuals are drawn and their phenotypes are simulated by applying a multinomial model with baseline risks for two diseases of 0.1 and 0.05 (e.g. AE and psoriasis) mimicking the respective prevalence using a previously described algorithm.[112] The controls were distributed equally on both case sets. The minor allele frequencies (MAF) $\in (0.1, 0.2, 0.3)$ and the odds ratios (OR) $\in (1.15, 1.2, 1.3)$ are varied. Power is estimated for levels of $\alpha = 0.001$ and $\alpha = 10^{-5}$ with R=1,000 replicates to detect (a) disease specific, (b) agonistic and (c) antagonistic effects. Disease specific effects are alway simulated for the first trait with baseline risk set to 0.1. For sensitivity analysis either an equal baseline risk for both diseases is set, or the controls are distributed proportionally on both case sets.
To compare type I error between CCMA and SBMA under the null hypothesis H_0 of no association between genetic markers and any disease, 20,000 replicates are simulated with the same settings as described above.

2.5. Multinomial Regression Model

The multinomial regression model (MNM) is used to validate CCMA on the pre-filtered SNV set and to estimate simultaneously SNV effects on both diseases. MNM is carried out with the `nnet`-package[113] in R.[111]
The MNM can be written as

$$P(Y_i = r) = \frac{exp(\beta_{r0} + \mathbf{x}_i^T \boldsymbol{\beta}_r)}{1 + \sum_{j=1}^{q-1} exp(\beta_{j0} + \mathbf{x}_i^T \boldsymbol{\beta}_j)}, \quad r = 1, \ldots, q-1 \tag{2.34}$$

where q is defined as the number of possible response categories, hence atopic eczema, psoriasis or healthy control. This can be equivalently written as

$$log \frac{P(Y_i = r)}{P(Y_i = q)} = \beta_{r0} + \mathbf{x}_i^T \boldsymbol{\beta}_r \tag{2.35}$$

with $Y_i = r$ as the r-th response category (atopic eczema or psoriasis) and $Y_i = q$ as the reference category, here healthy control. The vector $\mathbf{x}_i$ of independent variables, together with the coefficients $\boldsymbol{\beta}_r$, determines the log odds for category r with respect to the reference category q (see page 73 in Fahrmeir & Tutz[114]).
Here the response categories 1 = atopic eczema, 2 = psoriasis and 3 = controls are modeled and controls are set as reference category. In addition to the particular SNV, independent co-variables such as *age*, *sex* and the first four principal components derived from the MDS analysis (see Section A.3), which controls for population stratification, are used. The parameters are described in Table 2.2.

	Regression coefficients	
Explanatory variable	**atopic eczema**	**psoriasis**
intercept	β_{10}	β_{20}
sex	$\beta_{1\ \text{sex}}$	$\beta_{2\ \text{sex}}$
age	$\beta_{1\ \text{age}}$	$\beta_{2\ \text{age}}$
1st Principal Component	$\beta_{1\ \text{PC1}}$	$\beta_{2\ \text{PC1}}$
2nd Principal Component	$\beta_{1\ \text{PC2}}$	$\beta_{2\ \text{PC2}}$
3rd Principal Component	$\beta_{1\ \text{PC3}}$	$\beta_{2\ \text{PC3}}$
4th Principal Component	$\beta_{1\ \text{PC4}}$	$\beta_{2\ \text{PC4}}$
SNV	$\beta_{1\ \text{SNV}}$	$\beta_{2\ \text{SNV}}$

Table 2.2.: Parameter estimates from the multinomial regression model for each SNV

In order to calculate P-values for the SNV effect, linear hypothesis testing using the *Wald statistic* is performed:

$$\omega = (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\xi})^T [\mathbf{C}\mathbf{F}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{C}^T]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\xi}) \tag{2.36}$$

where the matrix $\mathbf{C}$ describes the contrasts of the SNV effects, $\mathbf{F}^{-1}(\hat{\boldsymbol{\beta}})$ the observed covariance matrix of $\boldsymbol{\beta}_{r,SNV}$ and $\boldsymbol{\xi}$ a vector reflecting H_0 which is here set to 0 for all linear hypotheses. The P-value can be calculated by

$$P = 1 - F_{df}(\omega)$$

with df = degrees of freedom defined by rank($\mathbf{C}$) and F_{df} being the corresponding distribution function of the χ^2 distribution with given df.
For SNV effects five linear hypotheses are defined: *overall*, *atopic eczema*, *psoriasis*, *agonistic* and *antagonistic*. The corresponding matrices $\mathbf{C}$ are given in Table 2.3.

Hypothesis			**Matrix $\mathbf{C}_{df\times 16}$**	**df**
			$\overbrace{\text{1,...,8}}\ \overbrace{\text{1,...,8}}$	
overall	$H_0 : \begin{pmatrix} \beta_{1\ \text{SNV}}+\beta_{2\ \text{SNV}} \\ \beta_{1\ \text{SNV}}-\beta_{2\ \text{SNV}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$H_1 : \begin{pmatrix} \beta_{1\ \text{SNV}}+\beta_{2\ \text{SNV}} \\ \beta_{1\ \text{SNV}}-\beta_{2\ \text{SNV}} \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \dots 0\ 1\ 0 \dots 0\ 1 \\ 0 \dots 0\ 1\ 0 \dots 0\ \text{-1} \end{pmatrix}$	2
AE	$H_0 : \beta_{1\ \text{SNV}} = 0$	$H_1 : \beta_{1\ \text{SNV}} \neq 0$	$\begin{pmatrix} 0 \dots 0\ 1\ 0 \dots 0\ 0 \end{pmatrix}$	1
Psoriasis	$H_0 : \beta_{2\ \text{SNV}} = 0$	$H_1 : \beta_{2\ \text{SNV}} \neq 0$	$\begin{pmatrix} 0 \dots 0\ 0\ 0 \dots 0\ 1 \end{pmatrix}$	1
agonistic	$H_0 : \beta_{1\ \text{SNV}} + \beta_{2\ \text{SNV}} = 0$	$H_1 : \beta_{1\ \text{SNV}} + \beta_{2\ \text{SNV}} \neq 0$	$\begin{pmatrix} 0 \dots 0\ 1\ 0 \dots 0\ 1 \end{pmatrix}$	1
antagonistic	$H_0 : \beta_{1\ \text{SNV}} - \beta_{2\ \text{SNV}} = 0$	$H_1 : \beta_{1\ \text{SNV}} - \beta_{2\ \text{SNV}} \neq 0$	$\begin{pmatrix} 0 \dots 0\ 1\ 0 \dots 0\ \text{-1} \end{pmatrix}$	1

Table 2.3.: Definitions of matrix $\mathbf{C}$ for linear hypotheses. The indices 1 and 2 denote the response atopic eczema and psoriasis, respectively.

In analogy to the CCMA, the effects can be categorized according to the hypothesis that leads to the smallest P-value

$$\min(P_{\text{atopic eczema}}, P_{\text{psoriasis}}, P_{\text{agonistic}}, P_{\text{antagonistic}}) \tag{2.37}$$

which allows for a comparison of both methods.

2.6. COMBINED & OVERLAP Method

The COMBINED & OVERLAP method (COM) has been described by Ellinghaus et al..[6] For both diseases a genomewide meta-analysis is conducted as described in Section 2.1. In the OVERLAP approach previously established risk SNVs of one disease (A) will be checked for significance with a threshold of $P < 0.01$ in the meta-analysis results of the other disease (B) and vice versa. In the COMBINED approach the authors suggest performing a meta-analysis for the combined phenotype, i.e. the case-control studies of both diseases are meta-analyzed together. Meta-analyzing the effects for the same allele investigates agonistic effects. By combining the effect of one allele for disease A and the effect of the alternative allele for disease B allows researchers to investigate antagonistic effects. To declare a SNV exhibiting a pleiotropic effect, they postulated a combined P-value criterion:

$$P_A < 0.05\ \&\ P_B < 0.05\ \&\ P_{\text{COMBINED}} < 10^{-4}$$

where P_A and P_B in our case translate into $P_{\text{atopic eczema}}$ and $P_{\text{psoriasis}}$, respectively.
In order to compare effect categorization of COM with CCMA and MNM, the SNV effects are categorized according to the hypothesis that leads to the smallest P-value using eq. (2.37). To this end, $P_{\text{atopic eczema}}$ and $P_{\text{psoriasis}}$ are derived from the disease specific studies and $P_{\text{agonistic}}$ and $P_{\text{antagonistic}}$ are identified as described above.

2.7. Comparison of Effect Categorization

For evaluating the ability of the CCMA method to find single disease or pleiotropic effects the categorization of the SNV effects is compared with the categorization derived by the MNM as a gold standard. Then it is evaluated whether the CCMA method is superior to the previously proposed COM method in identifying pleiotropy.
While the SBMA can clearly determine agonistic effects, it is not trivial to distinguish antagonistic from disease-specific effects without imposing additional arbitrary rules (e.g. setting significance thresholds for $\tilde{P}^{+}_{\text{DLM}}$ and $\tilde{P}^{-}_{\text{DLM}}$, which both contribute to the two-sided P-value $\tilde{P}^{(2)}_{\text{DLM}}$ (2.9)). Therefore the SBMA method is used to select the set of SNVs with suggestive association ($P < 10^{-5}$), and based on this set CCMA is compared with COM and MNM in terms of effect categorization.
Ideally, the MNM would have been chosen for selecting the SNV set. However, the computational burden on a genome-wide scale using the MNM is too expensive. Hence for a fair comparison of the CCMA with the MNM and the COM methods the independent and frequently applied SBMA is chosen for SNV selection. As described above the CCMA will be separately compared with the SBMA by simulation studies.
Each of the selected SNVs is categorized into AE-specific, psoriasis-specific, agonistic or antagonistic according to the hypothesis that leads to the maximum test statistic for CCMA eq. (2.12) or the minimum $P - value$ for MNM and COM eq. (2.37). In a contingency table (Table 2.4) the concordance in effect categorization between two methods is displayed and the overall concordance rate is calculated by dividing the sums of the diagonal elements (e_{ij}) by the number of SNVs under consideration:

$$\text{concordance rate} = \frac{\sum_{i=1}^{4} e_{ii}}{\sum_{i=1}^{4}\sum_{j=1}^{4} e_{ij}} \tag{2.38}$$

	Method A			
Method B	AE	Agonistic	Antagonistic	Psoriasis
AE	e_{11}	e_{12}	e_{13}	e_{14}
Agonistic	e_{21}	e_{22}	e_{23}	e_{24}
Antagonistic	e_{31}	e_{32}	e_{33}	e_{34}
Psoriasis	e_{41}	e_{42}	e_{43}	e_{44}

Table 2.4.: Concordance of effect categorization between two methods.

3. Results

3.1. Comparison of Power and Type I Error Rate between SBMA and CCMA

The simulation based power analysis compares the SBMA with the CCMA method. Besides varying the MAFs and the ORs, the influence of changes in baseline risk and distribution of controls on both case sets is investigated. (1) In the beginning the baseline risk for the two diseases is set to 0.1 and 0.05, respectively, and the controls are distributed equally on both case sets. Then three modifications are investigated: (2) the baseline risk of both diseases is set equally to 0.1, (3) the controls are proportionally distributed on both case sets, (4) the case/control ratio is 1:2, and (5) case/control ratio=1:2 and equal baseline risk of 0.1 (Table 3.1).

	Baseline risk	Control	Disease 1: range		Disease 2: range	
No	Disease 1/2	distribution	#cases	#controls	#cases	#controls
(1)	0.1/0.05	equally	627-902	3353-3524	229-469	3353-3524
(2)	0.1/0.1	equally	598-868	3168-3389	487-863	3167-3388
(3)	0.1/0.05	proportionally	633-913	4219-5319	251-465	1550-2652
(4)	0.1/0.05	ca/ctrl ratio=1:2	1629-2172	3258-4344	672-1092	1344-2184
(5)	0.1/0.1	ca/ctrl ratio=1:2	1577-2093	3154-4186	1320-2044	2640-4088

Table 3.1.: Settings of the power analysis settings.

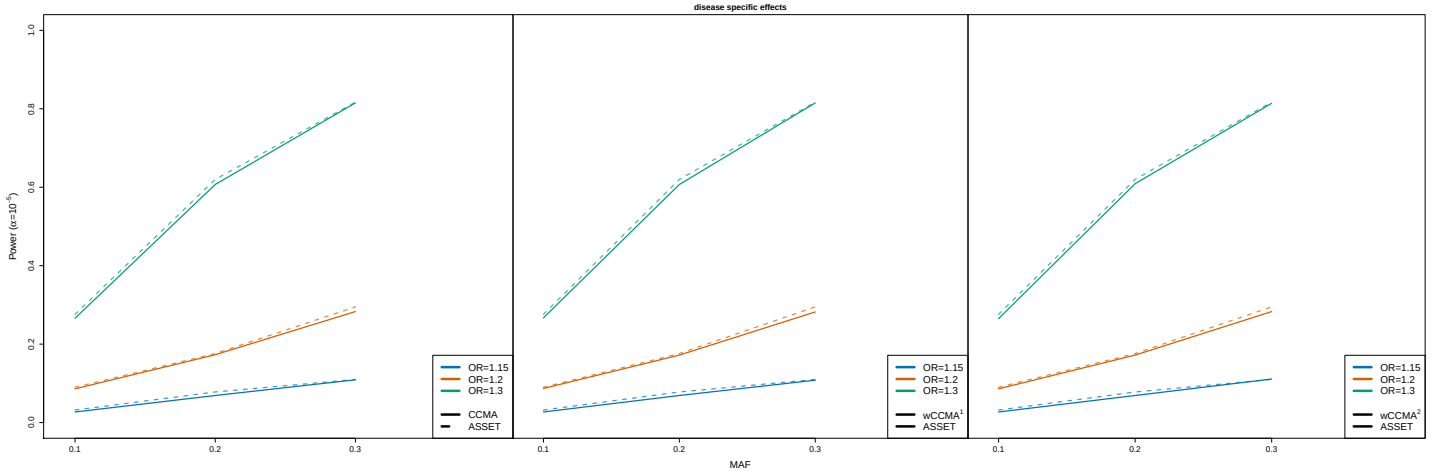
The simulation-based power analysis reveals that for detecting disease specific effects the CCMA has comparable power to the SBMA method at a significance level of 0.001, however it shows slightly increased power at a significance level of 10^{-5} (Figures 3.1, 3.4, 3.7, 3.10, 3.13).
For detecting agonistic effects the CCMA method is marginally less powerful than the SBMA method (Figures 3.2, 3.5, 3.8, 3.11, 3.14). In settings (2) and (5) with equal baseline risks for both diseases (Table 3.1), all three CCMA versions (CCMA, wCCMA^1, wCCMA^2) show very similar power. For all other settings wCCMA^1, the weighted version optimized for agonistic effects, is the most powerful CCMA test statistic.
For detecting antagonistic effects the CCMA is marginally less powerful than SBMA in settings (1) and (2) with equally distributed controls (Figures 3.3, 3.6). Of note, the difference is smallest for setting (2) at a significance level of 10^{-5}. For all other settings, the CCMA has (very) similar power to the SBMA. For the significance level of 10^{-5} in settings (3) and (4) with unequal baseline risks for the two diseases the wCCMA^2 is marginally more powerful for detecting antagonistic effects than SBMA (Figures 3.9,

3.12). In all settings the wCCMA^2, the weighted version optimized for antagonistic effects, is the most powerful CCMA test statistic.
In all settings, all three CCMA statistics (almost) hold the nominal significance level of 0.05 under the null hypothesis, when there is no effect (Tables 3.7, 3.8 and Figures 3.16, 3.17). In contrast, the SBMA shows increased type I error rates at $\alpha = 0.05$ compared to CCMA. For lower nominal significance levels ($\alpha \in 0.01, 0.005$) the type I error rate of CCMA is in some scenarios slightly above the thresholds, while the SBMA again shows higher type I error rates compared to CCMA. For $\alpha = 0.001$ a greater fluctuation of the type I error rate around the threshold is observed for both methods due to higher simulation variability. However, in most scenarios the SBMA reveals higher type I error rates than CCMA, which mostly is below the threshold.
In summary, the simulation-based power analysis reveals that the CCMA shows comparable power to the SBMA method for detecting disease specific or antagonistic effects and is marginally less powerful for detecting agonistic effects (Tables 3.2, 3.3, 3.4, 3.5, 3.6). However, CCMA provides better control of the type I error rate (Tables 3.7, 3.8, Figures 3.16, 3.17), which demonstrates the trade off between optimizing power and controlling type I error.

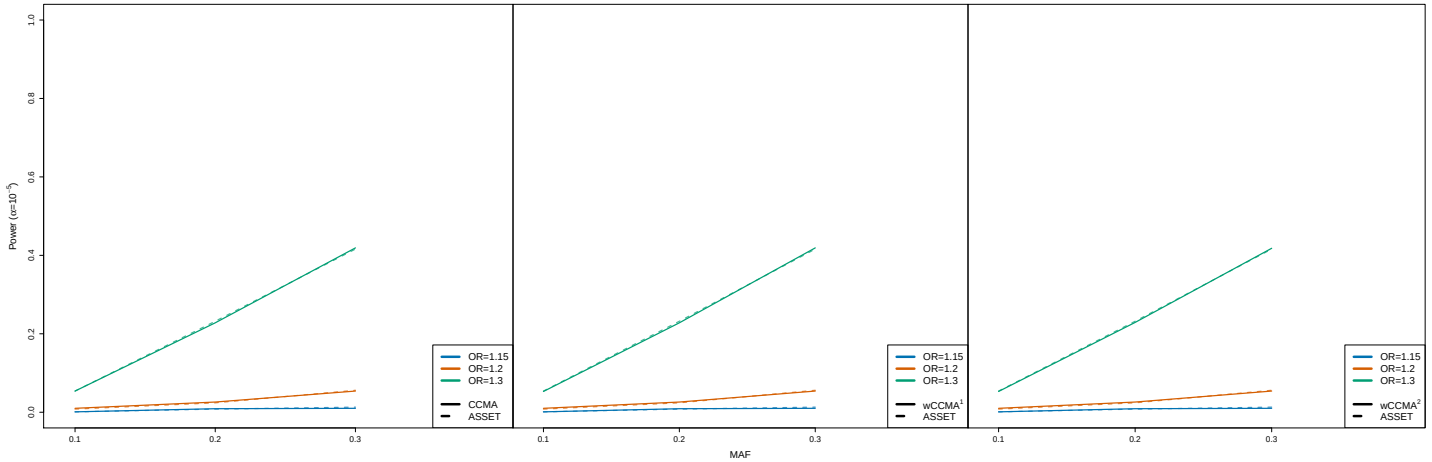
MAF	OR	disease-specific effect				agonistic effect				antagonistic effect			
		ASSET	CCMA	wCCMA[1]	wCCMA[2]	ASSET	CCMA	wCCMA[1]	wCCMA[2]	ASSET	CCMA	wCCMA[1]	wCCMA[2]
$\alpha = 0.001$													
0.1	1.15	0.0320	0.0270	0.0270	0.0270	0.0600	0.0520	0.0530	0.0510	0.0430	0.0360	0.0350	0.0360
	1.2	0.0900	0.0860	0.0870	0.0860	0.1620	0.1400	0.1420	0.1410	0.1140	0.1060	0.1050	0.1100
	1.3	0.2760	0.2660	0.2670	0.2650	0.5780	0.5420	0.5430	0.5400	0.4470	0.4330	0.4310	0.4350
0.2	1.15	0.0780	0.0690	0.0690	0.0690	0.1820	0.1700	0.1700	0.1710	0.1340	0.1300	0.1280	0.1300
	1.2	0.1760	0.1730	0.1720	0.1720	0.4430	0.4160	0.4190	0.4150	0.3450	0.3270	0.3250	0.3330
	1.3	0.6200	0.6070	0.6070	0.6090	0.9050	0.8920	0.8930	0.8900	0.8320	0.8200	0.8170	0.8220
0.3	1.15	0.1100	0.1090	0.1080	0.1110	0.2460	0.2240	0.2240	0.2270	0.2130	0.2000	0.1950	0.1980
	1.2	0.2950	0.2830	0.2820	0.2830	0.6130	0.5830	0.5870	0.5790	0.5330	0.5060	0.5040	0.5140
	1.3	0.8170	0.8150	0.8150	0.8140	0.9760	0.9670	0.9680	0.9660	0.9430	0.9360	0.9340	0.9390
$\alpha = 10^{-5}$													
0.1	1.15	0.0010	0.0010	0.0010	0.0010	0.0030	0.0020	0.0020	0.0020	0.0010	0.0020	0.0020	0.0020
	1.2	0.0080	0.0100	0.0100	0.0100	0.0220	0.0220	0.0220	0.0220	0.0140	0.0110	0.0110	0.0120
	1.3	0.0540	0.0540	0.0530	0.0530	0.1980	0.1880	0.1860	0.1870	0.0940	0.0910	0.0900	0.0910
0.2	1.15	0.0080	0.0090	0.0090	0.0090	0.0190	0.0190	0.0180	0.0200	0.0070	0.0070	0.0070	0.0070
	1.2	0.0240	0.0260	0.0260	0.0260	0.1010	0.0900	0.0910	0.0910	0.0630	0.0580	0.0570	0.0590
	1.3	0.2320	0.2280	0.2280	0.2290	0.5800	0.5540	0.5590	0.5520	0.4490	0.4210	0.4120	0.4290
0.3	1.15	0.0130	0.0100	0.0100	0.0100	0.0300	0.0260	0.0270	0.0250	0.0230	0.0240	0.0240	0.0230
	1.2	0.0560	0.0540	0.0540	0.0540	0.2090	0.1940	0.1950	0.1930	0.1380	0.1290	0.1250	0.1340
	1.3	0.4160	0.4190	0.4190	0.4180	0.8000	0.7830	0.7820	0.7810	0.6960	0.6790	0.6690	0.6830

[1]wCCMA using transformation matrix $\mathbf{A}_{(1)}$
[2]wCCMA using transformation matrix $\mathbf{A}_{(2)}$

Table 3.2.: **Setting 1.** Power comparison of the CCMA, wCCMA and Subset-Based Meta-Analysis (ASSET) for detection of true associations at a significance level of $\alpha = 0.001$ and $\alpha = 10^{-5}$. For each power estimate R=1,000 simulations with n=8,000 individuals for various MAF and OR values have been carried out. The disease status is assigned by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls equally distributed to both case sets.

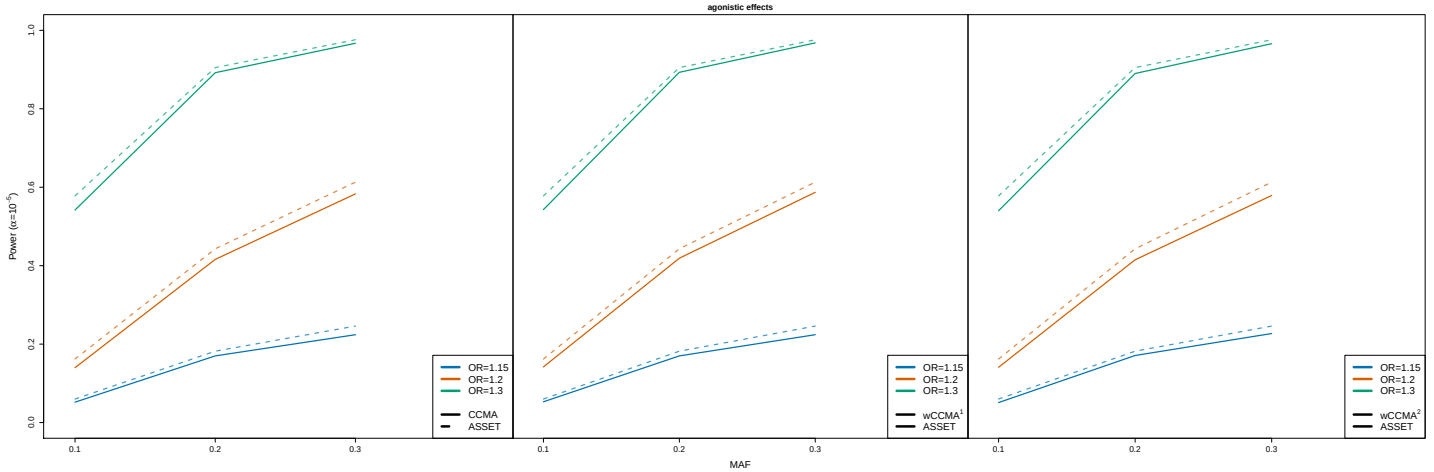


(a) $\alpha = 0.001$

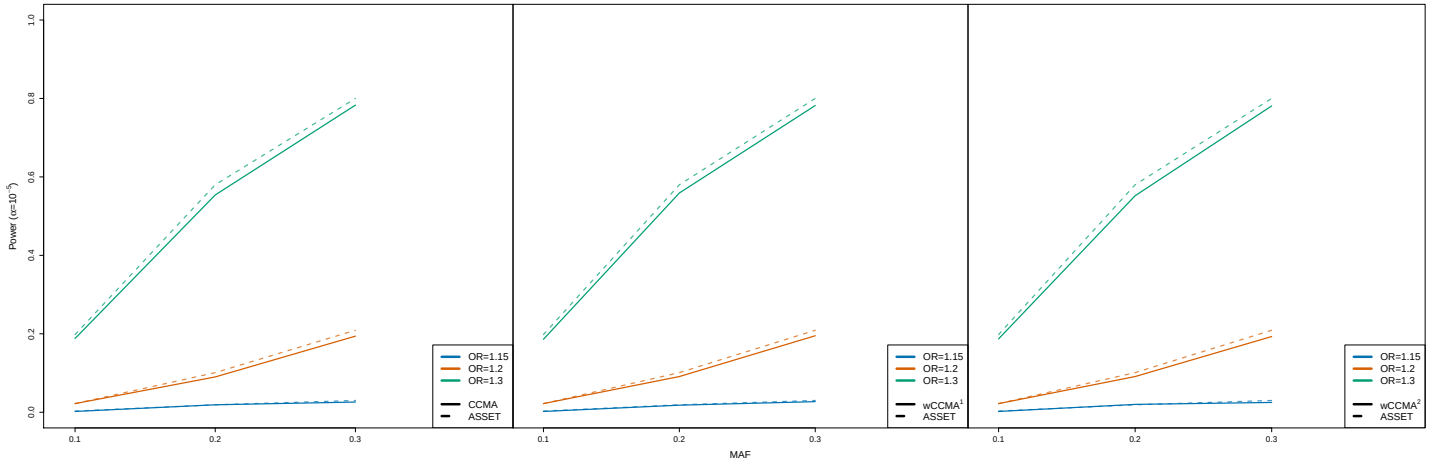


(b) $\alpha = 10^{-5}$

Figure 3.1.: **Setting 1: Disease specific effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting a disease specific effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls equally distributed to both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.

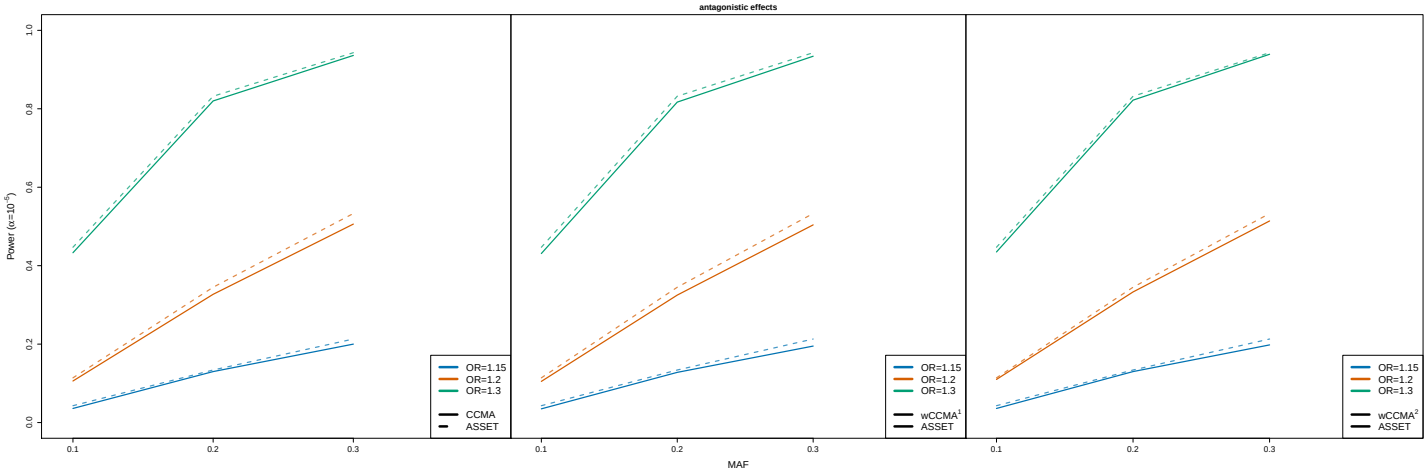


(a) $\alpha = 0.001$

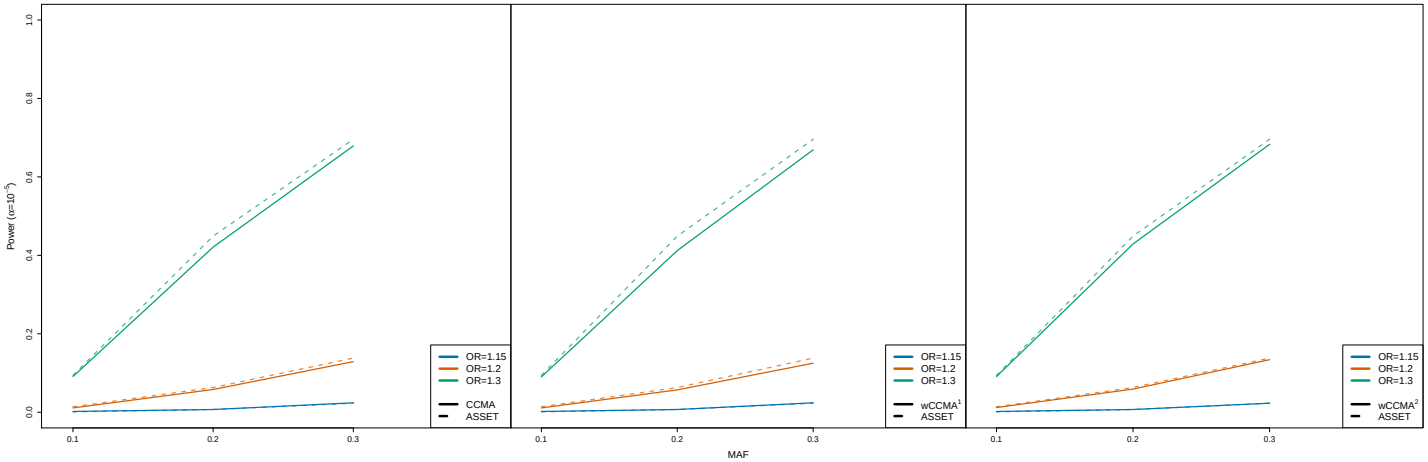


(b) $\alpha = 10^{-5}$

Figure 3.2.: **Setting 1: Agonistic effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting an agonistic effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls equally distributed to both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.



(a) $\alpha = 0.001$



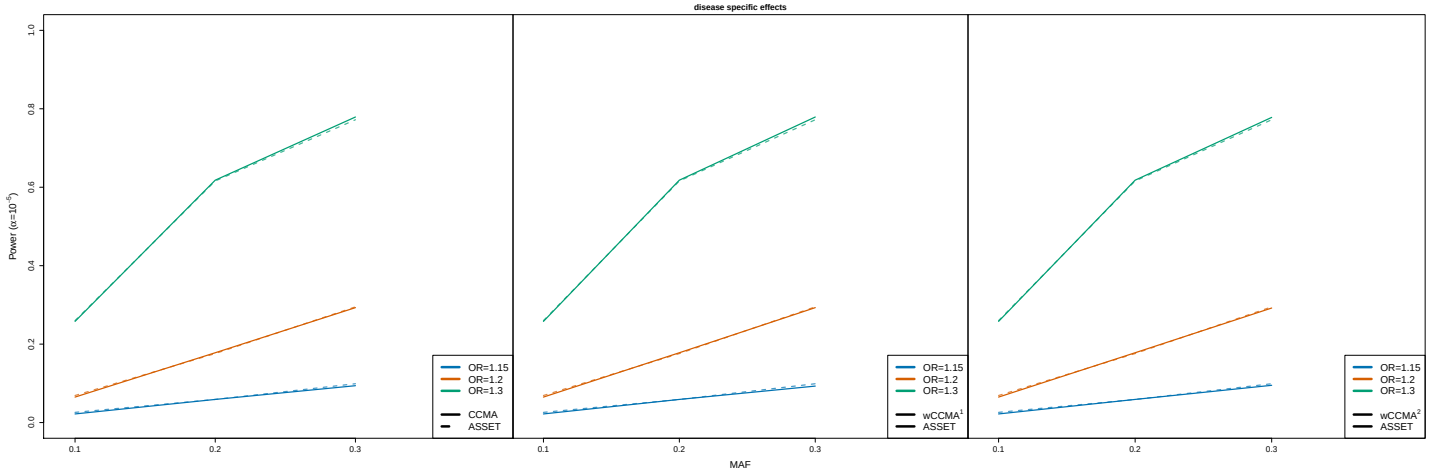
(b) $\alpha = 10^{-5}$

Figure 3.3.: **Setting 1: Antagonistic effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting an antagonistic effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls equally distributed to both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.

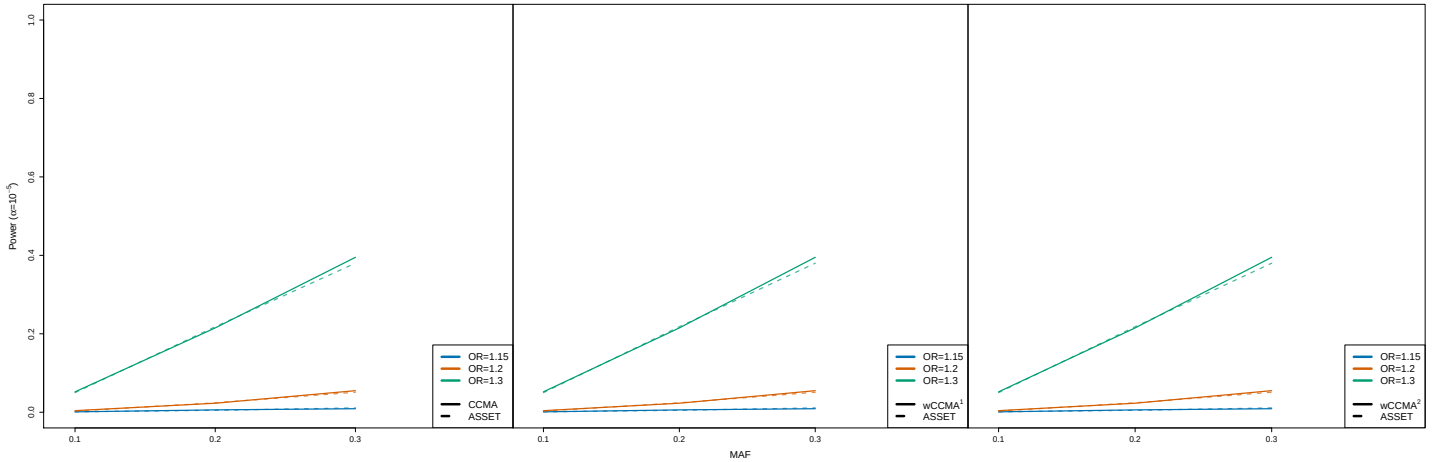
MAF	OR	disease-specific effect				agonistic effect				antagonistic effect			
		ASSET	CCMA	wCCMA[1]	wCCMA[2]	ASSET	CCMA	wCCMA[1]	wCCMA[2]	ASSET	CCMA	wCCMA[1]	wCCMA[2]
$\alpha = 0.001$													
0.1	1.15	0.0260	0.0220	0.0220	0.0220	0.0860	0.0810	0.0810	0.0810	0.0680	0.0640	0.0640	0.0630
	1.2	0.0690	0.0650	0.0650	0.0650	0.2340	0.2150	0.2150	0.2150	0.1680	0.1620	0.1620	0.1620
	1.3	0.2600	0.2580	0.2580	0.2580	0.6670	0.6430	0.6440	0.6430	0.5480	0.5340	0.5320	0.5360
0.2	1.15	0.0590	0.0590	0.0590	0.0590	0.2370	0.2220	0.2220	0.2220	0.1830	0.1710	0.1710	0.1710
	1.2	0.1760	0.1780	0.1780	0.1780	0.5420	0.5090	0.5090	0.5090	0.4470	0.4310	0.4330	0.4310
	1.3	0.6160	0.6180	0.6180	0.6180	0.9610	0.9530	0.9520	0.9530	0.9130	0.9070	0.9070	0.9070
0.3	1.15	0.0990	0.0940	0.0930	0.0950	0.3640	0.3390	0.3390	0.3400	0.3150	0.2970	0.2970	0.2970
	1.2	0.2950	0.2930	0.2930	0.2920	0.7610	0.7390	0.7390	0.7400	0.6540	0.6330	0.6310	0.6330
	1.3	0.7720	0.7790	0.7790	0.7780	0.9960	0.9960	0.9960	0.9960	0.9870	0.9850	0.9850	0.9850
$\alpha = 10^{-5}$													
0.1	1.15	0.0010	0.0010	0.0010	0.0010	0.0080	0.0070	0.0070	0.0070	0.0020	0.0020	0.0020	0.0020
	1.2	0.0030	0.0040	0.0040	0.0040	0.0330	0.0310	0.0310	0.0310	0.0190	0.0180	0.0180	0.0180
	1.3	0.0500	0.0520	0.0520	0.0520	0.2710	0.2600	0.2600	0.2600	0.1640	0.1530	0.1530	0.1530
0.2	1.15	0.0050	0.0060	0.0060	0.0060	0.0280	0.0250	0.0250	0.0250	0.0290	0.0260	0.0260	0.0260
	1.2	0.0240	0.0230	0.0230	0.0230	0.1670	0.1580	0.1580	0.1580	0.1190	0.1130	0.1130	0.1110
	1.3	0.2180	0.2150	0.2150	0.2150	0.7520	0.7390	0.7390	0.7390	0.6330	0.6320	0.6290	0.6330
0.3	1.15	0.0110	0.0090	0.0090	0.0090	0.0640	0.0570	0.0570	0.0570	0.0580	0.0510	0.0510	0.0520
	1.2	0.0510	0.0550	0.0550	0.0550	0.3350	0.3210	0.3210	0.3210	0.2460	0.2440	0.2450	0.2460
	1.3	0.3800	0.3950	0.3950	0.3950	0.9250	0.9200	0.9200	0.9200	0.8550	0.8440	0.8410	0.8440

[1]wCCMA using transformation matrix $\mathbf{A}_{(1)}$
[2]wCCMA using transformation matrix $\mathbf{A}_{(2)}$

Table 3.3.: **Setting 2.** Power comparison of the CCMA, wCCMA and Subset-Based Meta-Analysis (ASSET) for detection of true associations at a significance level of $\alpha = 0.001$ and $\alpha = 10^{-5}$. For each power estimate R=1,000 simulations with n=8,000 individuals for various MAF and OR values have been carried out. The disease status is assigned by a multinomial model with baseline risk of 0.1 for both diseases and controls equally distributed to both case sets.

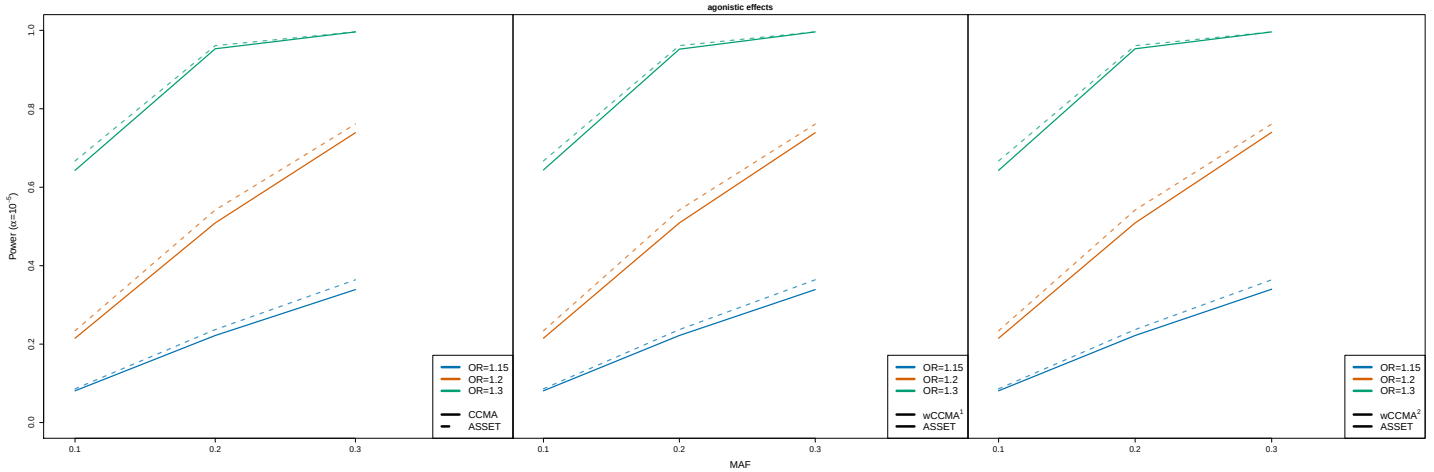


(a) $\alpha = 0.001$

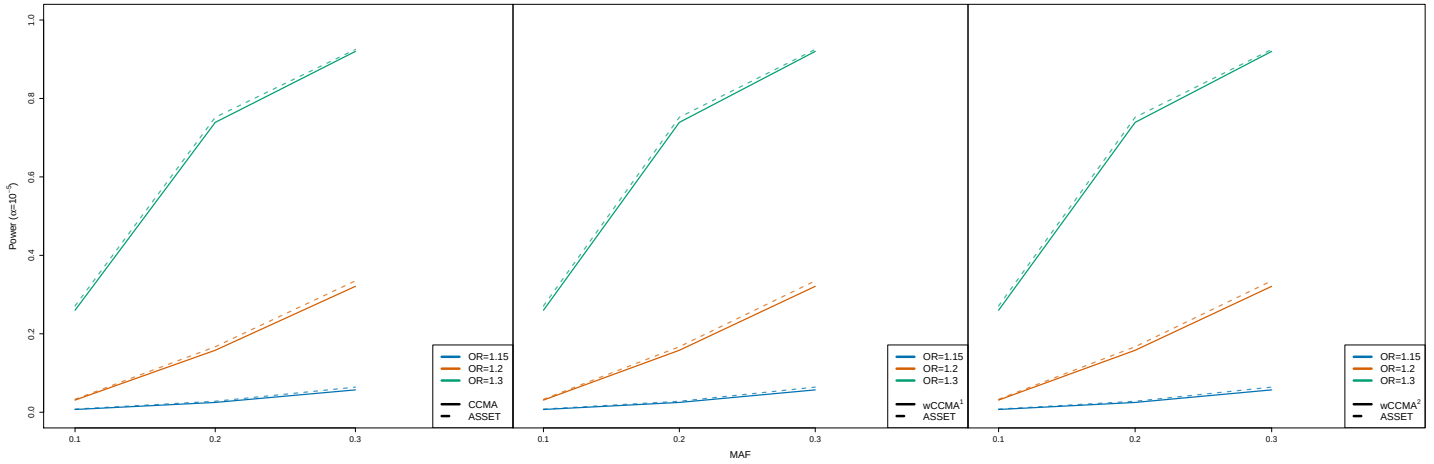


(b) $\alpha = 10^{-5}$

Figure 3.4.: **Setting 2: Disease specific effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting a disease specific effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risk of 0.1 for both diseases and controls equally distributed to both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.

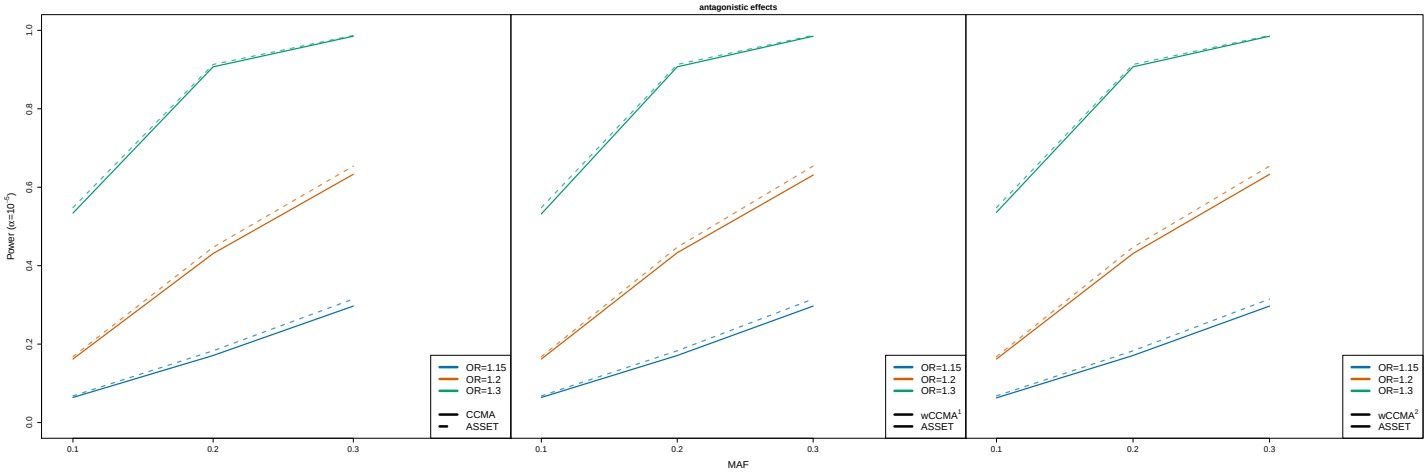


(a) $\alpha = 0.001$

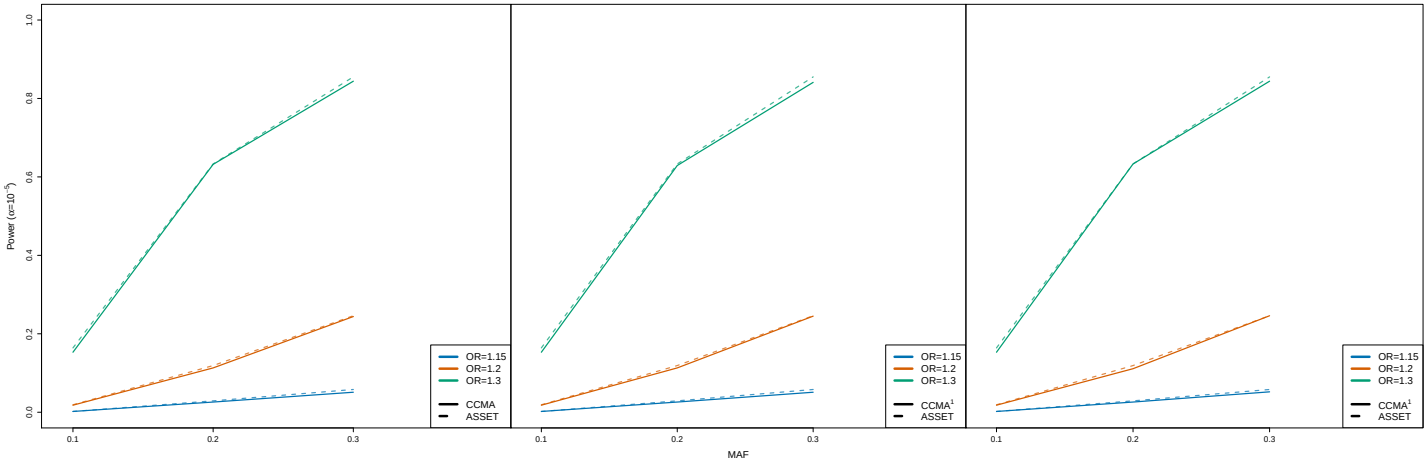


(b) $\alpha = 10^{-5}$

Figure 3.5.: **Setting 2: Agonistic effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting an agonistic effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risk of 0.1 for both diseases and controls equally distributed to both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.



(a) $\alpha = 0.001$



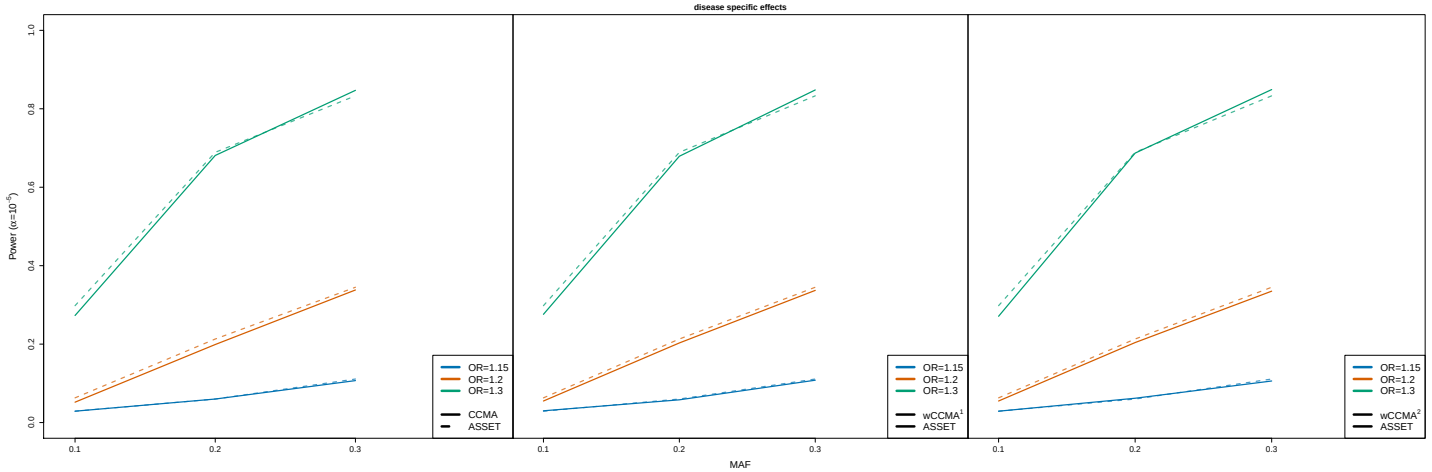
(b) $\alpha = 10^{-5}$

Figure 3.6.: **Setting 2: Antagonistic effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting an antagonistic effect. For each power estimate, we ran R=1,000 simulation with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risks for both diseases of 0.1 and controls equally distributed to both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.

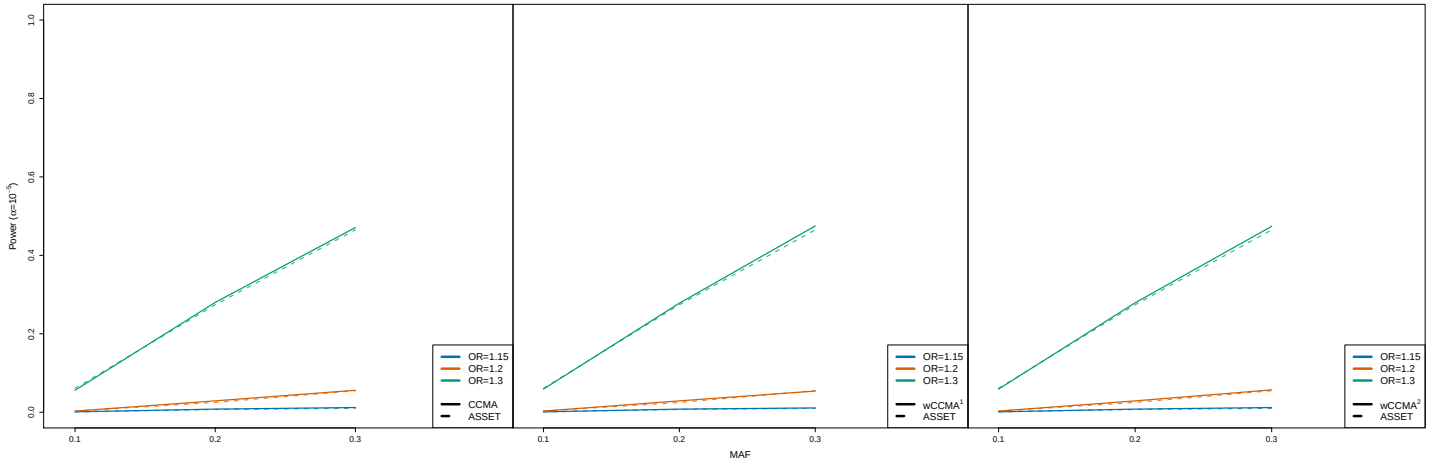
MAF	OR	disease-specific effect				agonistic effect				antagonistic effect			
		ASSET	CCMA	wCCMA[1]	wCCMA[2]	ASSET	CCMA	wCCMA[1]	wCCMA[2]	ASSET	CCMA	wCCMA[1]	wCCMA[2]
$\alpha = 0.001$													
0.1	1.15	0.0290	0.0290	0.0300	0.0290	0.0610	0.0570	0.0530	0.0530	0.0420	0.0450	0.0420	0.0420
	1.2	0.0630	0.0520	0.0550	0.0550	0.1910	0.1760	0.1730	0.1630	0.1350	0.1290	0.1110	0.1360
	1.3	0.2980	0.2730	0.2760	0.2710	0.5560	0.5190	0.5280	0.4870	0.4670	0.4490	0.4230	0.4590
0.2	1.15	0.0600	0.0600	0.0580	0.0620	0.1760	0.1590	0.1590	0.1450	0.1460	0.1320	0.1130	0.1350
	1.2	0.2130	0.1990	0.2030	0.2040	0.4120	0.3740	0.3770	0.3540	0.3340	0.3260	0.2950	0.3390
	1.3	0.6890	0.6810	0.6790	0.6870	0.9100	0.8980	0.9050	0.8840	0.8640	0.8540	0.8160	0.8550
0.3	1.15	0.1110	0.1070	0.1080	0.1060	0.2780	0.2470	0.2510	0.2280	0.2290	0.2160	0.1970	0.2220
	1.2	0.3450	0.3380	0.3370	0.3350	0.5980	0.5690	0.5680	0.5370	0.5100	0.4960	0.4610	0.5020
	1.3	0.8330	0.8470	0.8480	0.8490	0.9870	0.9820	0.9850	0.9760	0.9550	0.9490	0.9290	0.9530
$\alpha = 10^{-5}$													
0.1	1.15	0.0010	0.0010	0.0010	0.0010	0.0050	0.0040	0.0050	0.0030	0.0020	0.0020	0.0020	0.0020
	1.2	0.0030	0.0030	0.0030	0.0030	0.0290	0.0260	0.0270	0.0240	0.0160	0.0120	0.0070	0.0160
	1.3	0.0610	0.0560	0.0590	0.0590	0.1810	0.1710	0.1690	0.1490	0.1180	0.1010	0.0810	0.1250
0.2	1.15	0.0070	0.0080	0.0080	0.0080	0.0200	0.0180	0.0160	0.0150	0.0140	0.0130	0.0080	0.0140
	1.2	0.0250	0.0290	0.0290	0.0290	0.0910	0.0800	0.0820	0.0660	0.0710	0.0590	0.0520	0.0740
	1.3	0.2740	0.2800	0.2780	0.2790	0.6170	0.5880	0.5980	0.5440	0.4680	0.4520	0.3950	0.4780
0.3	1.15	0.0100	0.0120	0.0110	0.0120	0.0520	0.0450	0.0490	0.0350	0.0370	0.0350	0.0250	0.0390
	1.2	0.0550	0.0560	0.0540	0.0570	0.1880	0.1630	0.1720	0.1400	0.1460	0.1400	0.1100	0.1530
	1.3	0.4650	0.4710	0.4750	0.4740	0.8320	0.8060	0.8140	0.7640	0.7380	0.7190	0.6300	0.7430

[1]wCCMA using transformation matrix $\mathbf{A}_{(1)}$
[2]wCCMA using transformation matrix $\mathbf{A}_{(2)}$

Table 3.4.: **Setting 3.** Power comparison of the CCMA, wCCMA and Subset-Based Meta-Analysis (ASSET) for detection of true associations at a significance level of $\alpha = 0.001$ and $\alpha = 10^{-5}$. For each power estimate R=1,000 simulations with n=8,000 individuals for various MAF and OR values have been carried out. The disease status is assigned by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls proportionally distributed to both case sets.

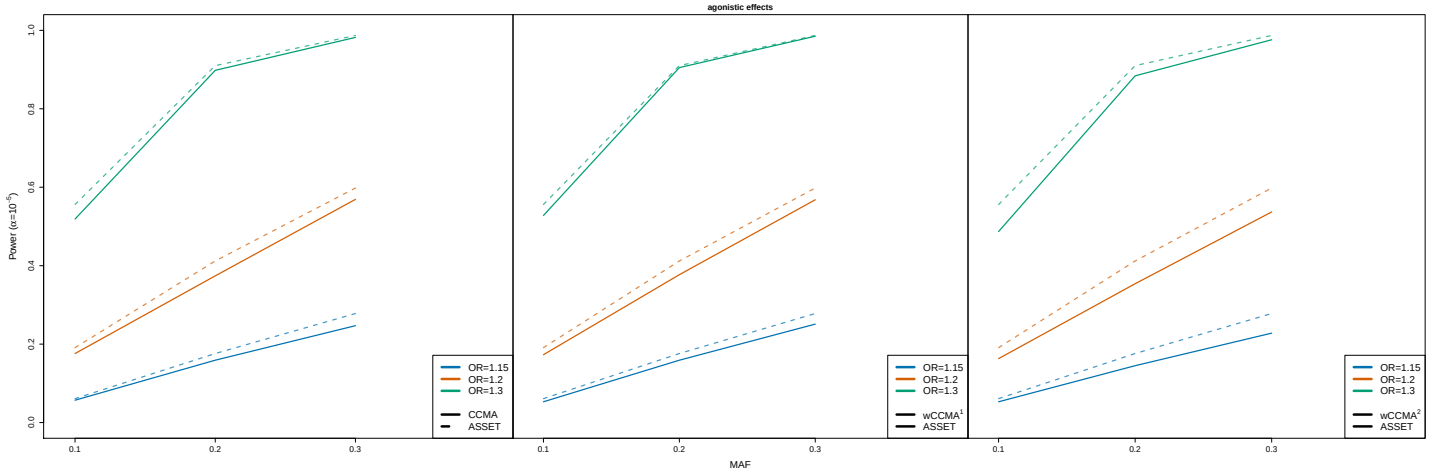


(a) $\alpha = 0.001$

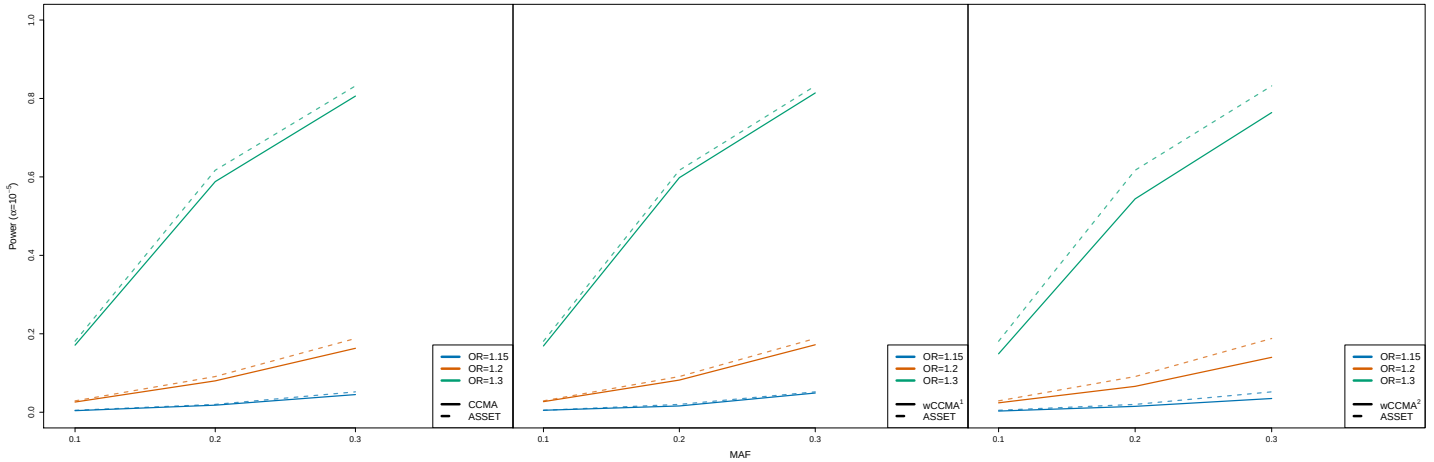


(b) $\alpha = 10^{-5}$

Figure 3.7.: **Setting 3: Disease specific effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting a disease specific effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls proportionally distributed to both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.

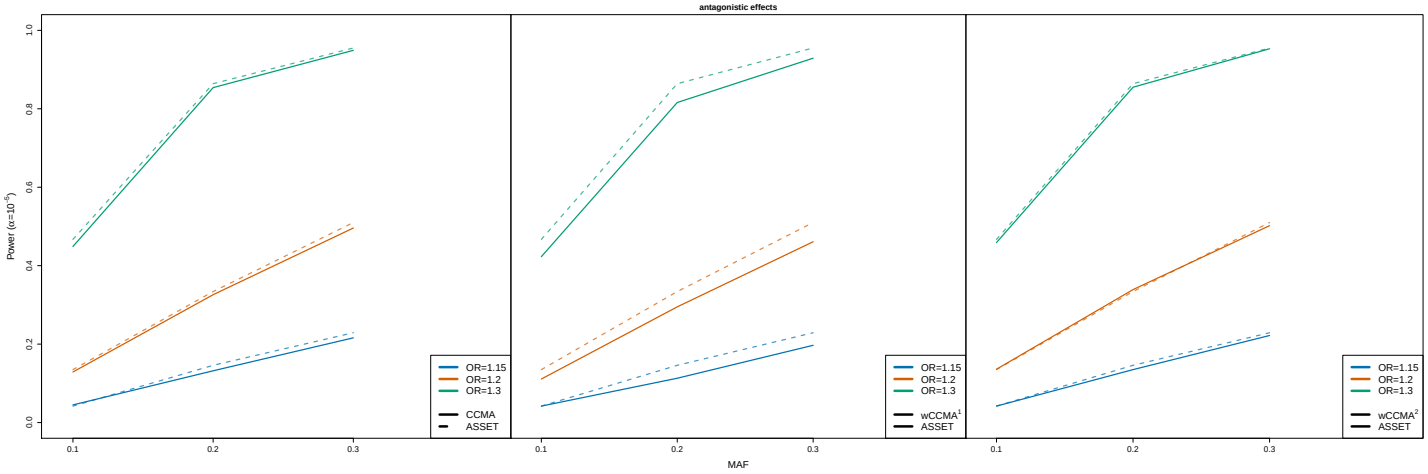


(a) $\alpha = 0.001$

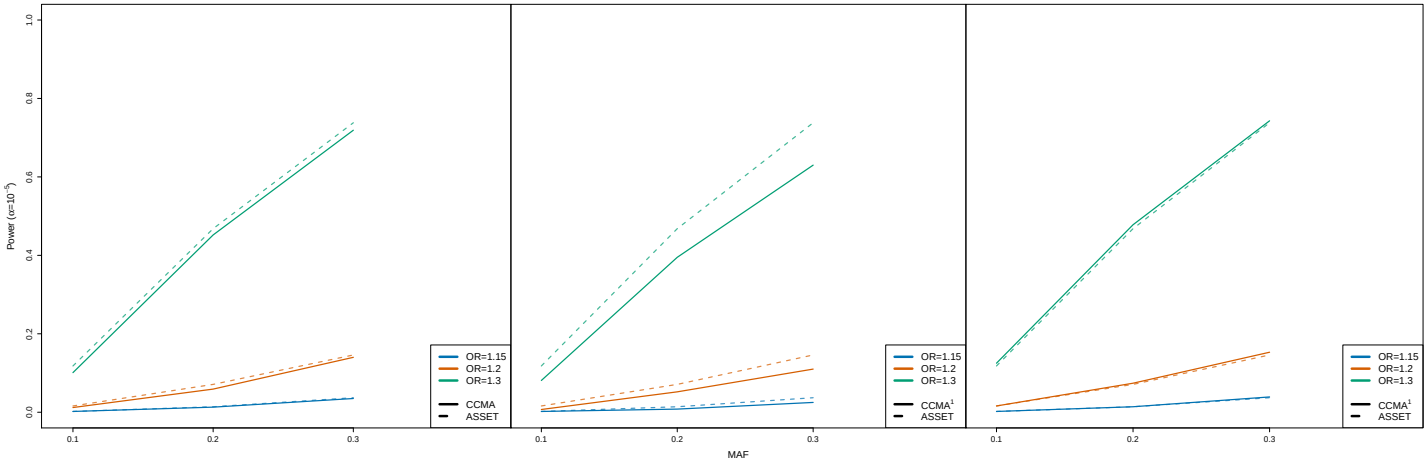


(b) $\alpha = 10^{-5}$

Figure 3.8.: **Setting 3: Agonistic effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting an agonistic effect. For each power estimate, we ran R,=1000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls proportionally distributed to both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.



(a) $\alpha = 0.001$



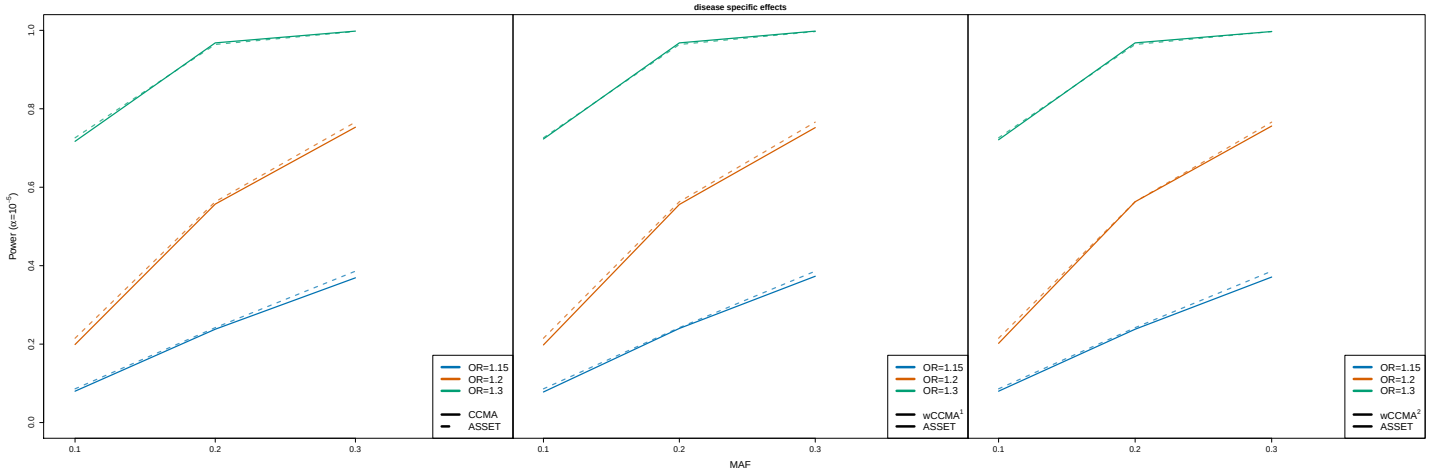
(b) $\alpha = 10^{-5}$

Figure 3.9.: **Setting 3: Antagonistic effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting an antagonistic effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls proportionally distributed to both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.

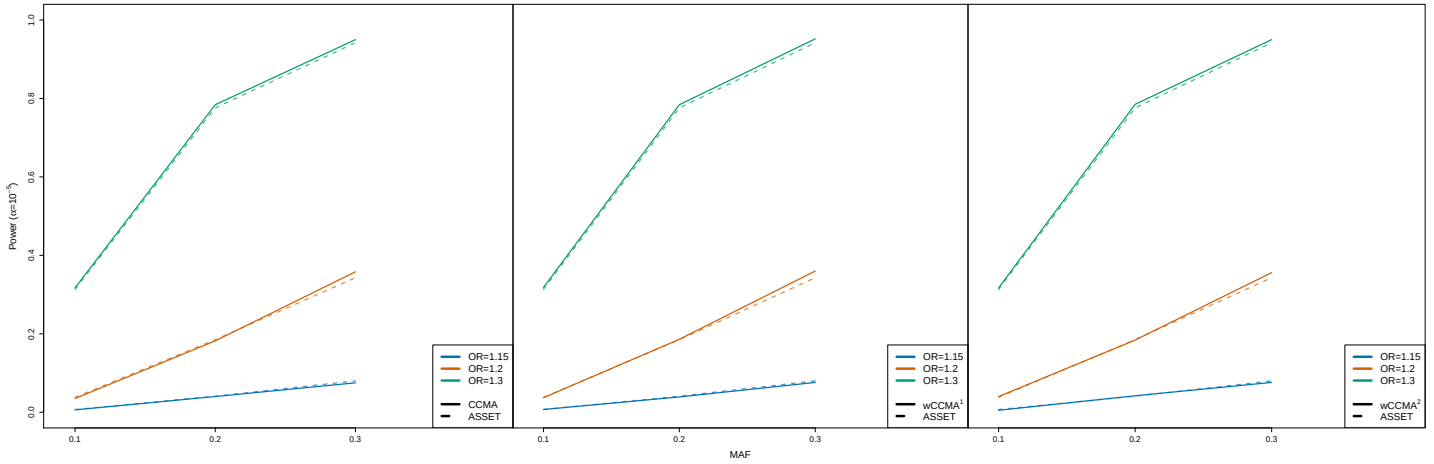
MAF	OR	disease-specific effect				agonistic effect				antagonistic effect			
		ASSET	CCMA	wCCMA[1]	wCCMA[2]	ASSET	CCMA	wCCMA[1]	wCCMA[2]	ASSET	CCMA	wCCMA[1]	wCCMA[2]
$\alpha = 0.001$													
0.1	1.15	0.0860	0.0800	0.0780	0.0800	0.1920	0.1730	0.1700	0.1590	0.1620	0.1600	0.1460	0.1580
	1.2	0.2150	0.1990	0.1980	0.2020	0.4720	0.4370	0.4490	0.4210	0.3750	0.3600	0.3380	0.3710
	1.3	0.7260	0.7170	0.7230	0.7210	0.9120	0.8990	0.9020	0.8880	0.8970	0.8880	0.8610	0.8960
0.2	1.15	0.2420	0.2380	0.2400	0.2380	0.4950	0.4790	0.4790	0.4500	0.4260	0.4090	0.3830	0.4150
	1.2	0.5630	0.5570	0.5560	0.5630	0.8550	0.8300	0.8320	0.8050	0.7820	0.7610	0.7320	0.7730
	1.3	0.9640	0.9680	0.9680	0.9680	0.9990	0.9990	0.9990	0.9980	0.9940	0.9940	0.9930	0.9940
0.3	1.15	0.3860	0.3690	0.3730	0.3710	0.6610	0.6350	0.6440	0.5970	0.6200	0.6020	0.5610	0.6150
	1.2	0.7660	0.7530	0.7520	0.7560	0.9560	0.9500	0.9510	0.9410	0.9380	0.9320	0.9120	0.9350
	1.3	0.9970	0.9980	0.9980	0.9970	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\alpha = 10^{-5}$													
0.1	1.15	0.0070	0.0060	0.0070	0.0050	0.0270	0.0230	0.0260	0.0190	0.0200	0.0180	0.0130	0.0210
	1.2	0.0380	0.0350	0.0370	0.0400	0.1320	0.1110	0.1210	0.0940	0.0800	0.0740	0.0540	0.0780
	1.3	0.3130	0.3170	0.3180	0.3170	0.6410	0.6070	0.6210	0.5550	0.5310	0.5200	0.4440	0.5400
0.2	1.15	0.0410	0.0400	0.0390	0.0420	0.1270	0.1200	0.1160	0.0980	0.0970	0.0850	0.0670	0.0950
	1.2	0.1850	0.1820	0.1860	0.1840	0.4770	0.4610	0.4640	0.4110	0.3790	0.3620	0.2990	0.3800
	1.3	0.7760	0.7840	0.7840	0.7850	0.9820	0.9780	0.9790	0.9710	0.9550	0.9440	0.9180	0.9560
0.3	1.15	0.0800	0.0750	0.0760	0.0760	0.2400	0.2170	0.2230	0.1970	0.2070	0.1940	0.1560	0.2120
	1.2	0.3430	0.3580	0.3600	0.3560	0.7220	0.6990	0.7060	0.6510	0.6640	0.6260	0.5560	0.6670
	1.3	0.9420	0.9500	0.9520	0.9500	0.9990	0.9980	0.9990	0.9960	0.9970	0.9960	0.9920	0.9970

[1]wCCMA using transformation matrix $\mathbf{A}_{(1)}$
[2]wCCMA using transformation matrix $\mathbf{A}_{(2)}$

Table 3.5.: **Setting 4.** Power comparison of the CCMA, wCCMA and Subset-Based Meta-Analysis (ASSET) for detection of true associations at a significance level of $\alpha = 0.001$ and $\alpha = 10^{-5}$. For each power estimate R=1,000 simulations with n=8,000 individuals for various MAF and OR values have been carried out. The disease status is assigned by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls distributed by a case control ratio of 1:2 for both case sets.

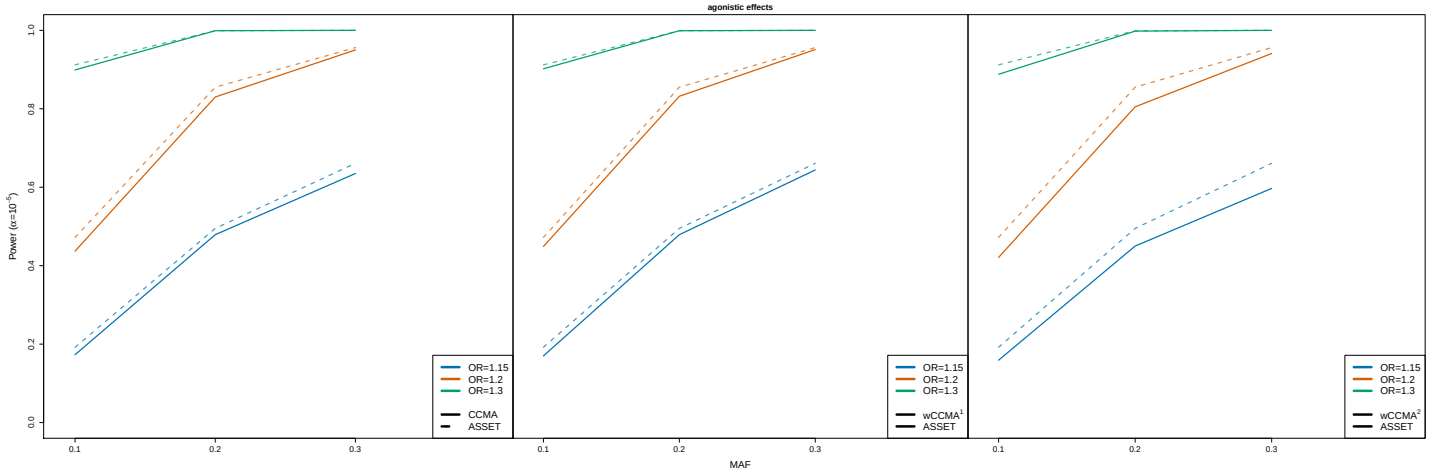


(a) $\alpha = 0.001$

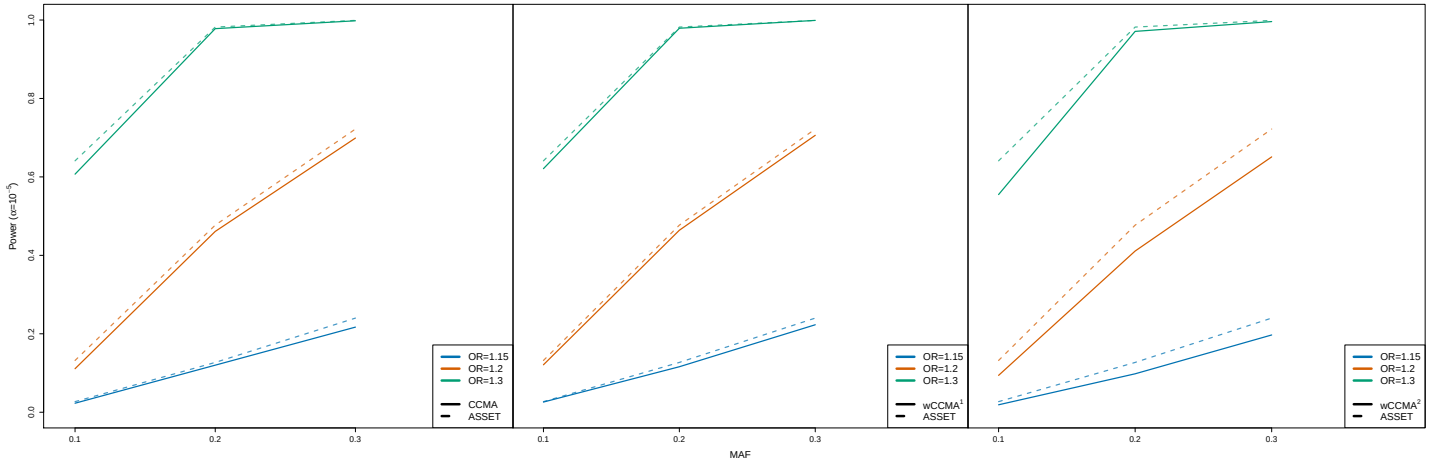


(b) $\alpha = 10^{-5}$

Figure 3.10.: **Setting 4: Disease specific effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting a disease specific effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls distributed by a case control ratio of 1:2 for both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.

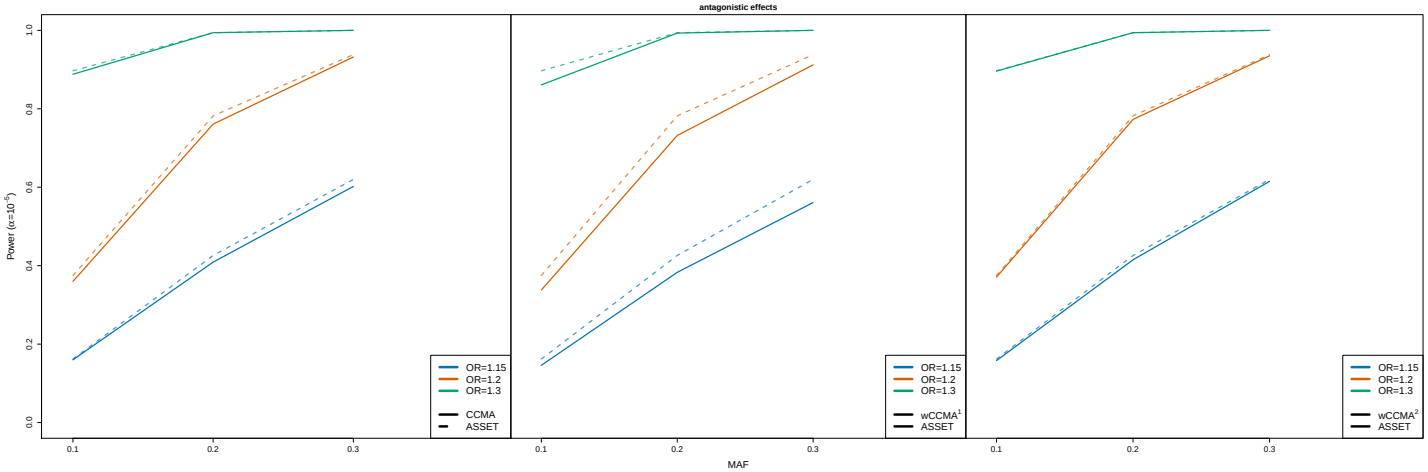


(a) $\alpha = 0.001$

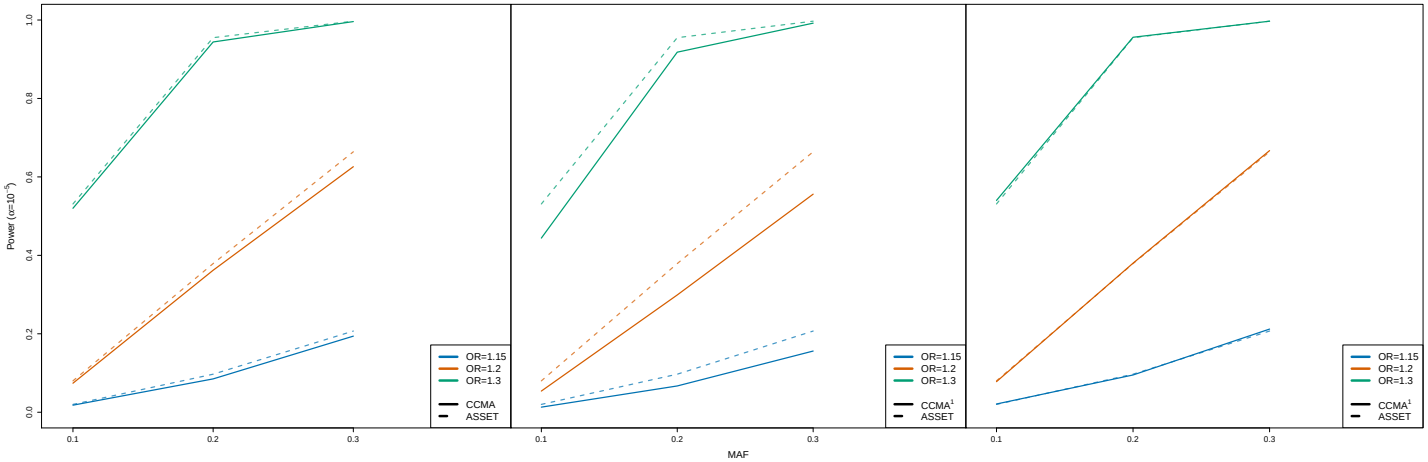


(b) $\alpha = 10^{-5}$

Figure 3.11.: **Setting 4: Agonistic effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting an agonistic effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls distributed by a case control ratio of 1:2 for both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.



(a) $\alpha = 0.001$



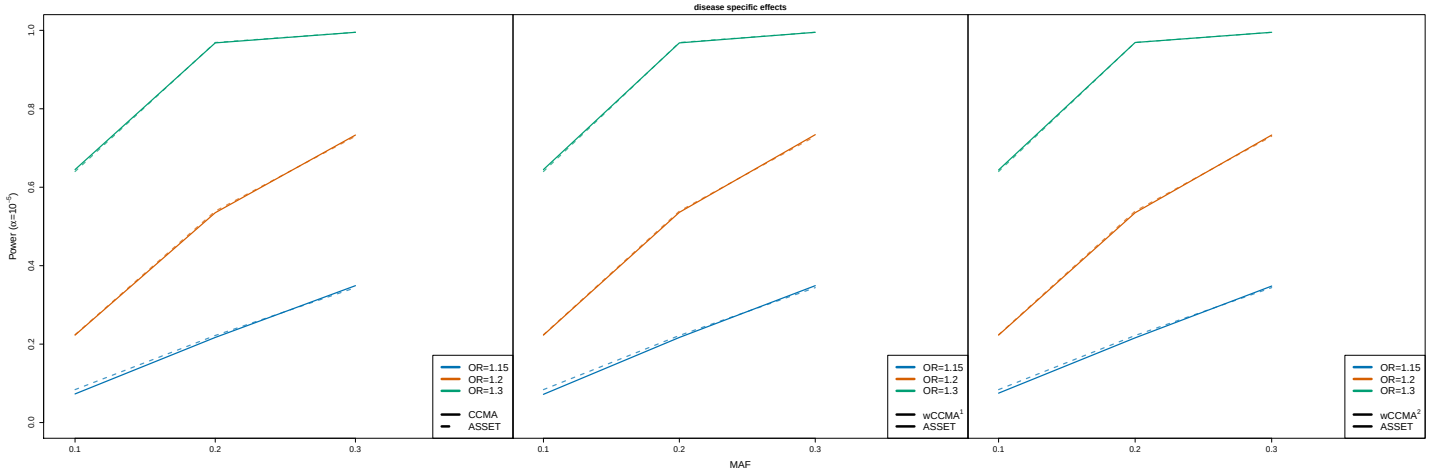
(b) $\alpha = 10^{-5}$

Figure 3.12.: **Setting 4: Antagonistic effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting an antagonistic effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risks of 0.1 and 0.05 for the two diseases, respectively, and controls distributed by a case control ratio of 1:2 for both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.

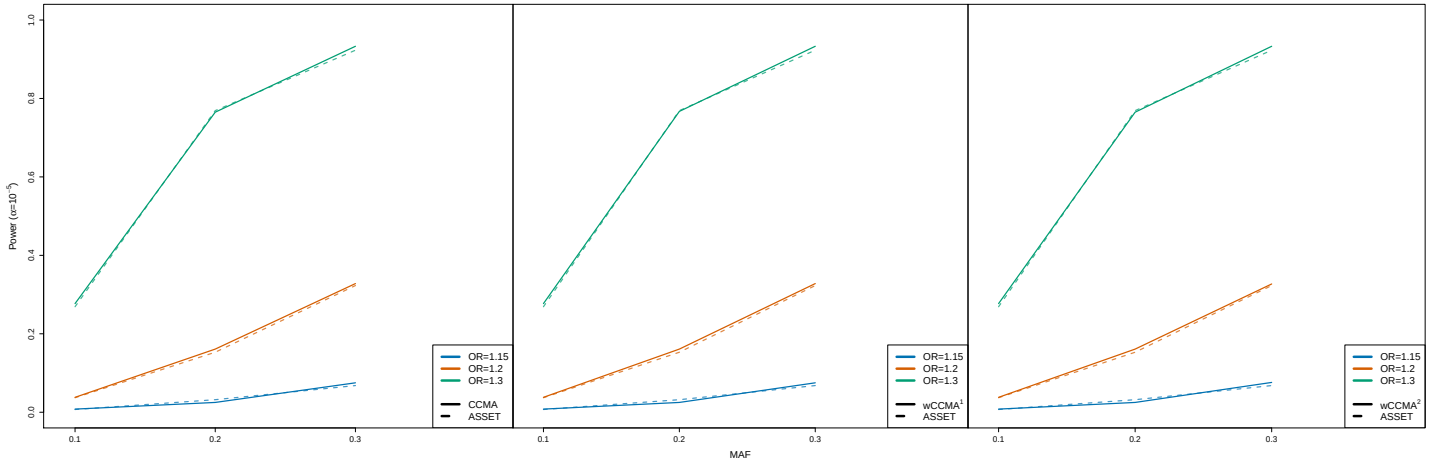
MAF	OR	disease-specific effect				agonistic effect				antagonistic effect			
		ASSET	CCMA	wCCMA[1]	wCCMA[2]	ASSET	CCMA	wCCMA[1]	wCCMA[2]	ASSET	CCMA	wCCMA[1]	wCCMA[2]
$\alpha = 0.001$													
0.1	1.15	0.0840	0.0730	0.0720	0.0750	0.2890	0.2600	0.2600	0.2600	0.2140	0.1970	0.1970	0.1970
	1.2	0.2240	0.2230	0.2230	0.2230	0.6060	0.5810	0.5840	0.5810	0.5250	0.5180	0.5170	0.5150
	1.3	0.6400	0.6450	0.6450	0.6440	0.9740	0.9690	0.9690	0.9690	0.9610	0.9560	0.9540	0.9570
0.2	1.15	0.2220	0.2170	0.2170	0.2160	0.6490	0.6240	0.6230	0.6230	0.5980	0.5830	0.5830	0.5850
	1.2	0.5390	0.5350	0.5360	0.5350	0.9350	0.9280	0.9290	0.9280	0.9230	0.9180	0.9170	0.9160
	1.3	0.9690	0.9680	0.9680	0.9690	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.3	1.15	0.3440	0.3490	0.3490	0.3480	0.8420	0.8250	0.8250	0.8230	0.8060	0.7860	0.7790	0.7870
	1.2	0.7300	0.7330	0.7340	0.7330	0.9890	0.9860	0.9860	0.9860	0.9850	0.9820	0.9800	0.9830
	1.3	0.9950	0.9950	0.9950	0.9950	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\alpha = 10^{-5}$													
0.1	1.15	0.0070	0.0080	0.0080	0.0080	0.0410	0.0400	0.0400	0.0390	0.0410	0.0370	0.0380	0.0370
	1.2	0.0370	0.0380	0.0380	0.0380	0.2140	0.2050	0.2050	0.2050	0.1420	0.1350	0.1320	0.1370
	1.3	0.2690	0.2770	0.2770	0.2770	0.8170	0.8020	0.8020	0.8030	0.7200	0.7010	0.7000	0.6980
0.2	1.15	0.0320	0.0250	0.0250	0.0250	0.2240	0.2110	0.2120	0.2110	0.1770	0.1780	0.1810	0.1780
	1.2	0.1530	0.1610	0.1610	0.1610	0.6680	0.6520	0.6520	0.6510	0.6180	0.6100	0.6060	0.6110
	1.3	0.7690	0.7650	0.7670	0.7650	0.9990	0.9990	0.9990	0.9990	0.9850	0.9840	0.9840	0.9850
0.3	1.15	0.0680	0.0750	0.0750	0.0760	0.4540	0.4320	0.4320	0.4340	0.3670	0.3540	0.3500	0.3590
	1.2	0.3230	0.3280	0.3280	0.3270	0.9110	0.9040	0.9040	0.9040	0.8200	0.8100	0.8020	0.8090
	1.3	0.9230	0.9330	0.9330	0.9330	1.0000	1.0000	1.0000	1.0000	0.9990	0.9990	1.0000	0.9990

[1]wCCMA using transformation matrix $\mathbf{A}_{(1)}$
[2]wCCMA using transformation matrix $\mathbf{A}_{(2)}$

Table 3.6.: **Setting 5.** Power comparison of the CCMA, wCCMA and Subset-Based Meta-Analysis (ASSET) for detection of true associations at a significance level of $\alpha = 0.001$ and $\alpha = 10^{-5}$. For each power estimate R=1,000 simulations with n=8,000 individuals for various MAF and OR values have been carried out. The disease status is assigned by a multinomial model with baseline risks of 0.1 for both diseases and controls distributed by a case control ratio of 1:2 for both case sets.

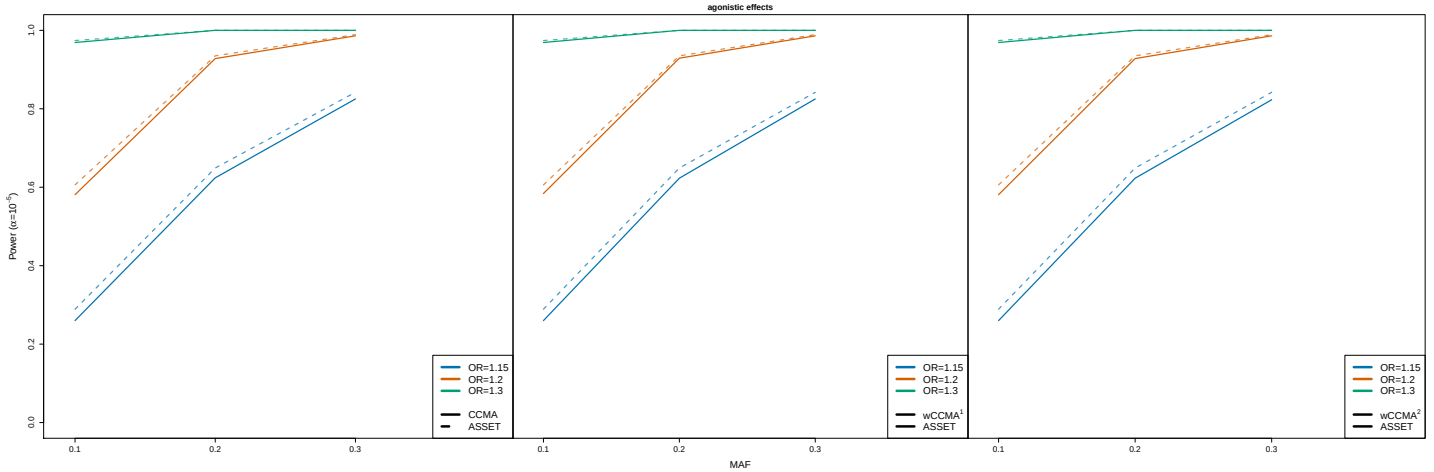


(a) $\alpha = 0.001$

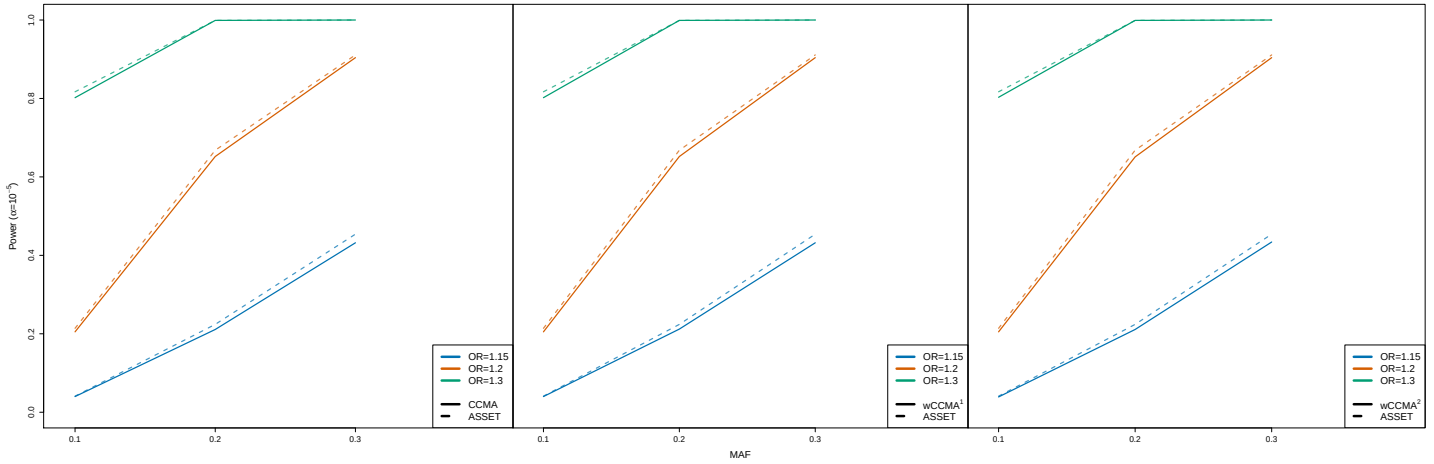


(b) $\alpha = 10^{-5}$

Figure 3.13.: **Setting 5: Disease specific effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting a disease specific effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risk of 0.1 for both diseases and controls distributed by a case control ratio of 1:2 for both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.

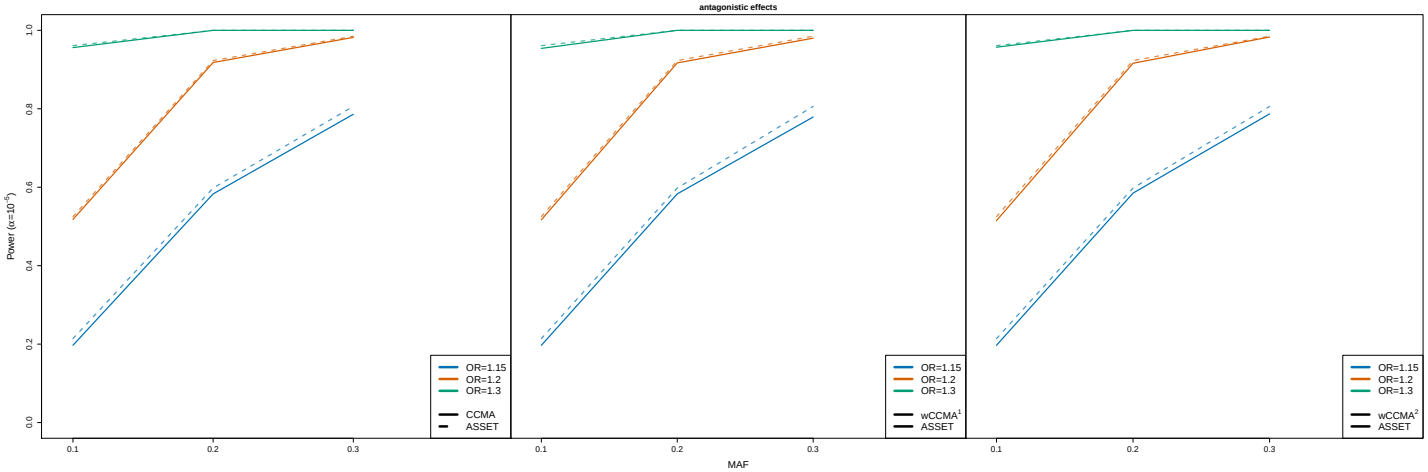


(a) $\alpha = 0.001$

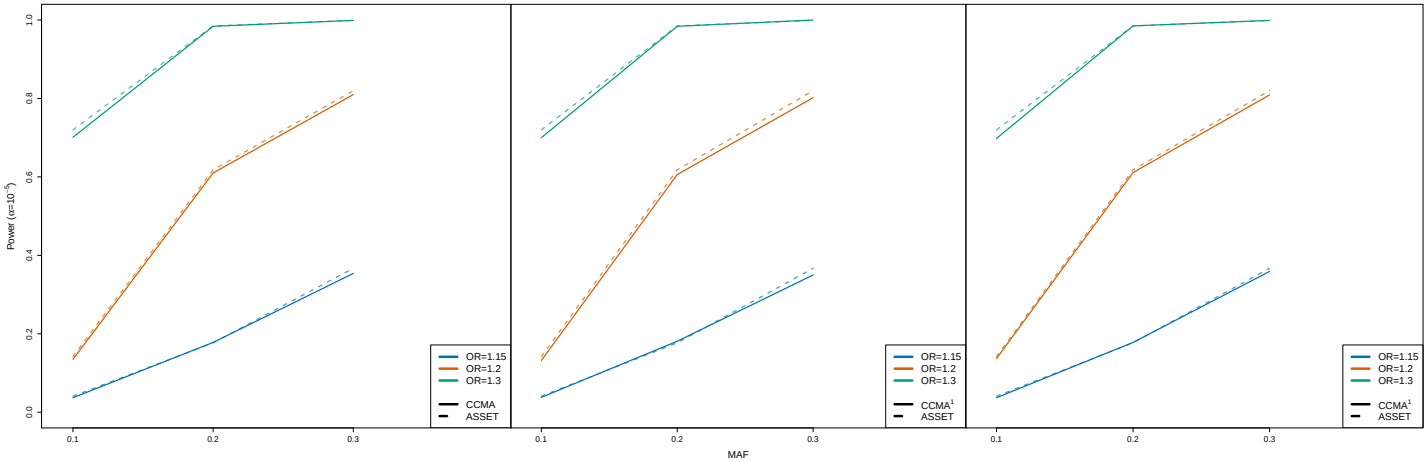


(b) $\alpha = 10^{-5}$

Figure 3.14.: **Setting 1: Agonistic effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting an agonistic effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risk of 0.1 for both diseases and controls distributed by a case control ratio of 1:2 for both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.



(a) $\alpha = 0.001$

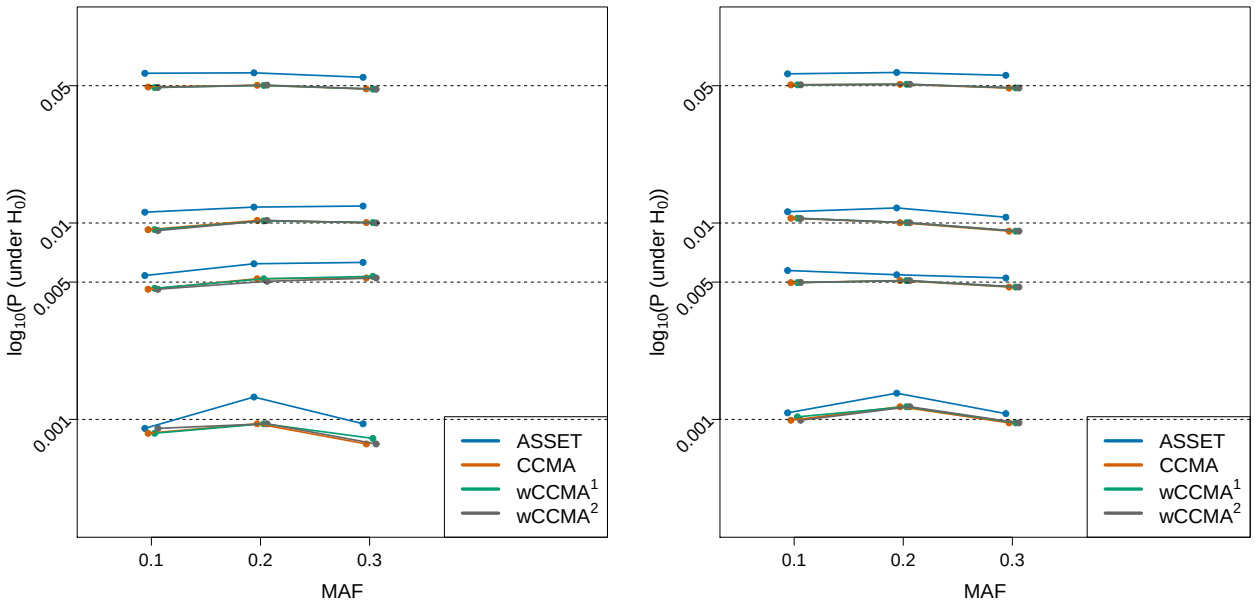


(b) $\alpha = 10^{-5}$

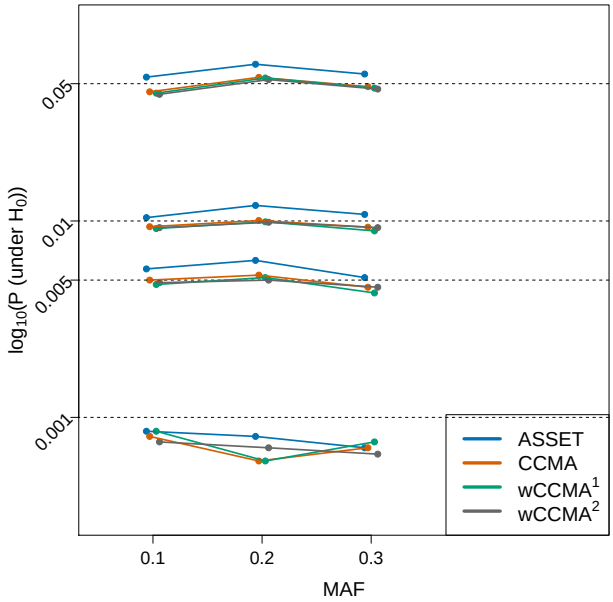
Figure 3.15.: **Setting 5: Antagonistic effect.** Simulation-based power comparison of CCMA and Subset-Based Meta-Analysis (ASSET) for detecting an antagonistic effect. For each power estimate, we ran R=1,000 simulations with n=8,000 individuals for various MAF and OR values and assigned the disease status by a multinomial model with baseline risk of 0.1 for both diseases and controls distributed by a case control ratio of 1:2 for both case sets. A significance threshold of $\alpha = 0.001$ and $\alpha = 10^{-5}$ was applied.

MAF	α	ASSET	CCMA	wCCMA[1]	wCCMA[2]
(1) unequal baseline risks; equally distributed controls					
0.1	0.001	0.00090	0.00085	0.00085	0.00090
	0.005	0.00540	0.00460	0.00465	0.00460
	0.01	0.01135	0.00925	0.00925	0.00915
	0.05	0.05785	0.04925	0.04895	0.04900
0.2	0.001	0.00130	0.00095	0.00095	0.00095
	0.005	0.00620	0.00520	0.00520	0.00505
	0.01	0.01205	0.01030	0.01025	0.01030
	0.05	0.05815	0.05025	0.05020	0.05035
0.3	0.001	0.00095	0.00075	0.00080	0.00075
	0.005	0.00630	0.00525	0.00535	0.00525
	0.01	0.01220	0.01005	0.01005	0.01000
	0.05	0.05520	0.04815	0.04805	0.04800
(2) equal baseline risk; equally distributed controls					
0.1	0.001	0.00108	0.00099	0.00103	0.00099
	0.005	0.00573	0.00497	0.00497	0.00498
	0.01	0.01141	0.01058	0.01058	0.01054
	0.05	0.05746	0.05052	0.05051	0.05053
0.2	0.001	0.00136	0.00116	0.00116	0.00116
	0.005	0.00545	0.00509	0.00509	0.00509
	0.01	0.01193	0.01004	0.01004	0.01003
	0.05	0.05840	0.05083	0.05087	0.05084
0.3	0.001	0.00107	0.00096	0.00096	0.00096
	0.005	0.00525	0.00472	0.00472	0.00472
	0.01	0.01070	0.00909	0.00909	0.00909
	0.05	0.05648	0.04863	0.04873	0.04866
(3) unequal baseline risks; proportionally distributed controls					
0.1	0.001	0.00085	0.00080	0.00085	0.00075
	0.005	0.00570	0.00500	0.00475	0.00485
	0.01	0.01040	0.00935	0.00915	0.00925
	0.05	0.05400	0.04545	0.04475	0.04415
0.2	0.001	0.00080	0.00060	0.00060	0.00070
	0.005	0.00630	0.00530	0.00515	0.00500
	0.01	0.01200	0.01005	0.00990	0.00985
	0.05	0.06280	0.05385	0.05345	0.05245
0.3	0.001	0.00070	0.00070	0.00075	0.00065
	0.005	0.00515	0.00460	0.00430	0.00460
	0.01	0.01080	0.00930	0.00890	0.00925
	0.05	0.05600	0.04830	0.04750	0.04695

Table 3.7.: **Settings 1-3.** Comparison of type I error rates for the CCMA, wCCMA and Subset-Based Meta-Analysis (ASSET) at nominal significance levels. For each scenario R=100,000 simulations with n=8,000 individuals for various MAF values have been carried out.



(1) (2)

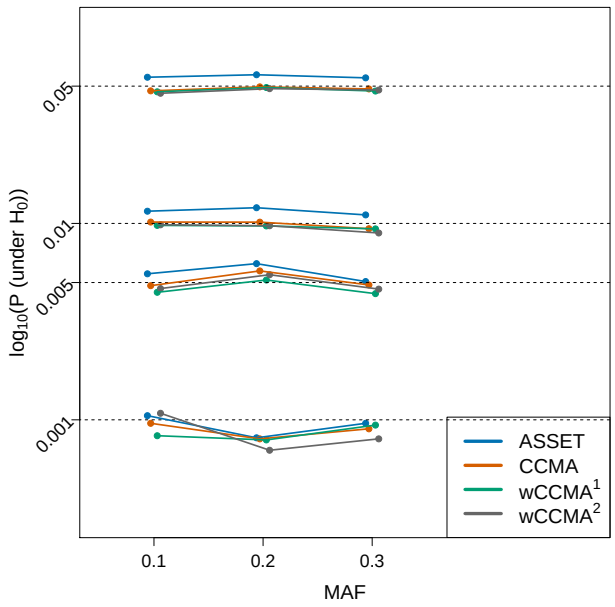


(3)

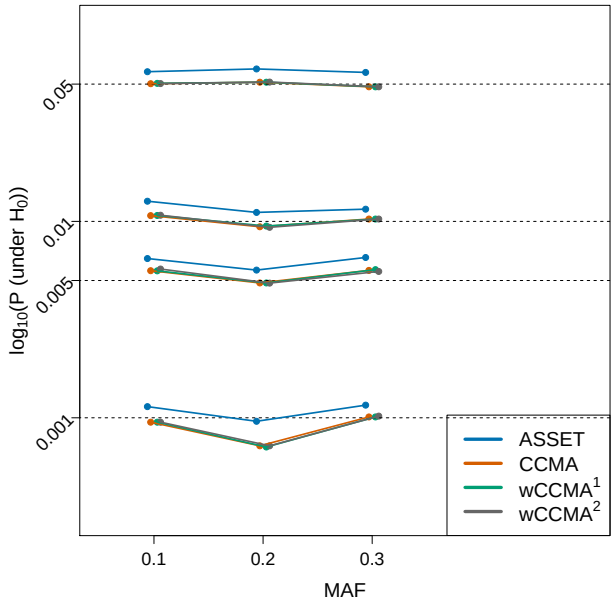
Figure 3.16.: **Settings 1-3.** Comparison of type I error rates for the CCMA, wCCMA and Subset-Based Meta-Analysis (ASSET) at nominal significance levels. For each scenario R=100,000 simulations with n=8,000 individuals for various MAF values have been carried out. (1) Unequal baseline risks of 0.1 and 0.05 for the two diseases, respectively, equally distributed controls; (2) equal baseline risk of 0.1 for both diseases, equally distributed controls; (3) unequal baseline risks of 0.1 and 0.05 for the two diseases, respectively, proportionally distributed controls.

MAF	α	ASSET	CCMA	wCCMA[1]	wCCMA[2]
(4) unequal baseline risks; case/control ratio=1:2					
0.1	0.001	0.00105	0.00096	0.00083	0.00108
	0.005	0.00554	0.00482	0.00446	0.00466
	0.01	0.01154	0.01016	0.00977	0.00984
	0.05	0.05550	0.04731	0.04679	0.04599
0.2	0.001	0.00081	0.00080	0.00079	0.00070
	0.005	0.00624	0.00573	0.00515	0.00548
	0.01	0.01202	0.01015	0.00971	0.00973
	0.05	0.05714	0.04958	0.04919	0.04854
0.3	0.001	0.00096	0.00090	0.00094	0.00080
	0.005	0.00506	0.00486	0.00439	0.00463
	0.01	0.01105	0.00940	0.00938	0.00894
	0.05	0.05511	0.04839	0.04721	0.04777
(5) equal baseline risk; case/control ratio=1:2					
0.1	0.001	0.00114	0.00095	0.00095	0.00095
	0.005	0.00647	0.00561	0.00560	0.00572
	0.01	0.01267	0.01070	0.01072	0.01075
	0.05	0.05782	0.05024	0.05045	0.05031
0.2	0.001	0.00096	0.00072	0.00071	0.00072
	0.005	0.00565	0.00486	0.00486	0.00485
	0.01	0.01111	0.00939	0.00947	0.00933
	0.05	0.05973	0.05113	0.05113	0.05121
0.3	0.001	0.00116	0.00101	0.00101	0.00102
	0.005	0.00655	0.00563	0.00568	0.00556
	0.01	0.01154	0.01028	0.01028	0.01028
	0.05	0.05735	0.04846	0.04843	0.04846

Table 3.8.: **Settings 4-5.** Comparison of type I error rates for the CCMA, wCCMA and Subset-Based Meta-Analysis (ASSET) at nominal significance levels. For each scenario R=100,000 simulations with n=8,000 individuals for various MAF values have been carried out.



(4)



(5)

Figure 3.17.: **Settings 4-5.** Comparison of type I error rates for the CCMA, wCCMA and Subset-Based Meta-Analysis (ASSET) at nominal significance levels. For each scenario R=100,000 simulations with n=8,000 individuals for various MAF values have been carried out. (4) Unequal baseline risks of 0.1 and 0.05 for the two diseases, respectively, case-control ratio=1:2; (5) equal baseline risk of 0.1 for both diseases, case-control ratio=1:2.

3.2. Assessment of Confidence Intervals

The data generated for the power comparison described in the previous Section 3.1 were also used to assess confidence intervals (a) defined for the CCMA test statistic $T_{\max}$, eq. (2.29), and (b) for the effect of interest (disease specific, agonistic or antagonistic), eqs. (2.31, 2.33).

3.2.1. Confidence Intervals for $T_{\max}$

Comparing the power of the CCMA test statistic at a nominal significance level of $\alpha = 0.05$ and the proportion of 95% confidence interval estimates > 0 shows identical results, if the alternative hypothesis is true (Tables 3.9, 3.10, Figures 3.18-3.23) emphasizing the duality of hypothesis test and confidence interval.
The confidence intervals are symmetric by construction using eq. (2.29), which can be seen in Figures 3.18-3.23.
The obtained confidence intervals for the test statistic $T_{\max}$ cannot be interpreted in terms of effect size, and thus no corresponding information can be derived. However, the confidence interval can be solved for the effect of interest (disease specific, agonistic or antagonistic) by eqs. $(2.31, 2.33)$, depending on which of the four statistics gives the maximum. This aspect is presented and discussed in the following subsection.

MAF	OR	disease-specific effect $P < 0.05$ 95%CI>0	agonistic effect $P < 0.05$ 95%CI>0	antagonistic effect $P < 0.05$ 95%CI>0
(1) unequal baseline risks; equally distributed controls				
0.1	1.15	0.271	0.367	0.338
	1.2	0.400	0.589	0.569
	1.3	0.525	0.702	0.690
0.2	1.15	0.436	0.581	0.531
	1.2	0.665	0.847	0.801
	1.3	0.772	0.923	0.909
0.3	1.15	0.723	0.910	0.884
	1.2	0.938	0.998	0.987
	1.3	0.986	1.000	1.000
(2) equal baseline risk; equally distributed controls				
0.1	1.15	0.252	0.447	0.405
	1.2	0.414	0.665	0.633
	1.3	0.504	0.798	0.770
0.2	1.15	0.398	0.695	0.633
	1.2	0.632	0.911	0.869
	1.3	0.759	0.978	0.949
0.3	1.15	0.705	0.958	0.903
	1.2	0.929	0.998	0.997
	1.3	0.971	0.999	0.999
(3) unequal baseline risks; proportionally distributed controls				
0.1	1.15	0.262	0.382	0.355
	1.2	0.425	0.628	0.561
	1.3	0.551	0.748	0.717
0.2	1.15	0.410	0.616	0.553
	1.2	0.687	0.831	0.800
	1.3	0.770	0.931	0.907
0.3	1.15	0.763	0.909	0.883
	1.2	0.955	0.992	0.990
	1.3	0.986	1.000	1.000

Table 3.9.: **Settings 1-3.** Power of the CCMA hypothesis test and the confidence interval estimates, both yield identical results. For each comparison R=1,000 simulations with n=8,000 individuals for various MAF and OR values have been carried out. The disease status is assigned by a multinomial model. (1) and (3): unequal baseline risks of 0.1 and 0.05 for the two diseases, respectively; (2): equal baseline risk of 0.1 for both diseases.

MAF	OR	disease-specific effect $P < 0.05$ 95%CI>0	agonistic effect $P < 0.05$ 95%CI>0	antagonistic effect $P < 0.05$ 95%CI>0
(4) unequal baseline risks; case/control ratio=1:2				
0.1	1.15	0.463	0.612	0.633
	1.2	0.721	0.874	0.862
	1.3	0.808	0.961	0.946
0.2	1.15	0.692	0.881	0.840
	1.2	0.912	0.978	0.979
	1.3	0.978	1.000	0.997
0.3	1.15	0.975	0.991	0.993
	1.2	0.997	1.000	1.000
	1.3	1.000	1.000	1.000
(5) equal baseline risk; case/control ratio=1:2				
0.1	1.15	0.428	0.750	0.689
	1.2	0.680	0.947	0.940
	1.3	0.833	0.984	0.978
0.2	1.15	0.695	0.934	0.926
	1.2	0.903	1.000	0.996
	1.3	0.973	1.000	0.999
0.3	1.15	0.943	0.999	0.997
	1.2	0.999	1.000	1.000
	1.3	1.000	1.000	1.000

Table 3.10.: **Settings 4-5.** Power of the CCMA hypothesis test and the confidence interval estimates, both yield identical results. For each comparison R=1,000 simulations with n=8,000 individuals for various MAF and OR values have been carried out. The disease status is assigned by a multinomial model. (4): unequal baseline risks of 0.1 and 0.05 for the two diseases, respectively; (5): equal baseline risk of 0.1 for both diseases.

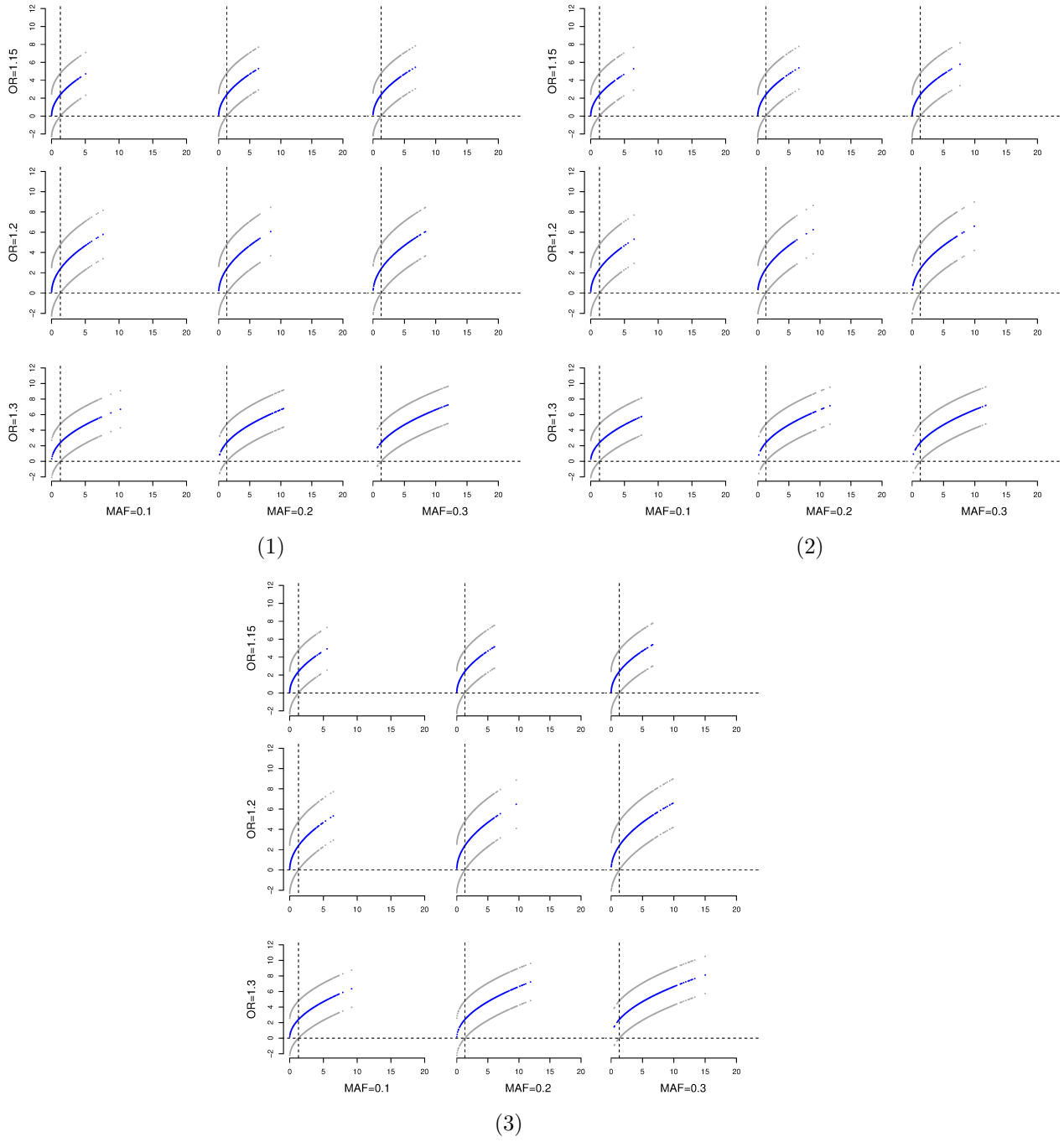
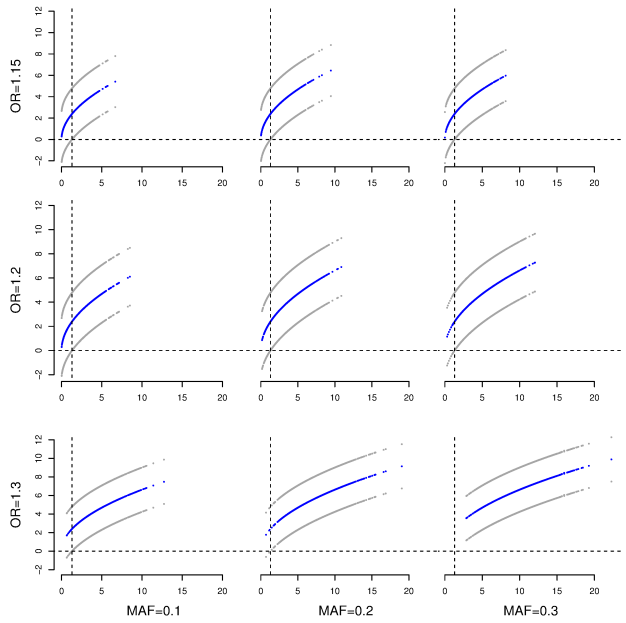
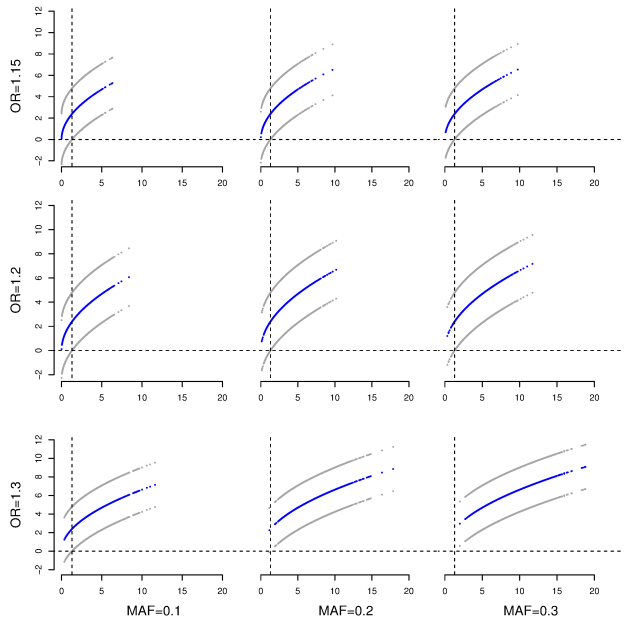


Figure 3.18.: **Settings 1-3. Disease specific effect.** Simulation-based comparison of $-log_{10}(P)$ (x axis) vs. T_{max} and 95% CI (y axis) for detecting disease specific effects with $P < 0.05$. Blue dots depict $T_{\max}$ and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (1) unequal baseline risks for the two diseases of 0.1 and 0.05, respectively, equally distributed controls; (2) equal baseline risk of 0.1 for both diseases, equally distributed controls; (3) unequal baseline risks for the two diseases of 0.1 and 0.05, respectively, proportionally distributed controls.



(4)



(5)

Figure 3.19.: **Settings 4-5. Disease specific effect.** Simulation-based comparison of $-log_{10}(P)$ (x axis) vs. T_{max} and 95% CI (y axis) for detecting disease specific effects with $P < 0.05$. Blue dots depict $T_{\max}$ and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (4) unequal baseline risks for the two diseases of 0.1 and 0.05, respectively, case-control ratio=1:2; (5) equal baseline risk of 0.1 for both diseases, case-control ratio=1:2.

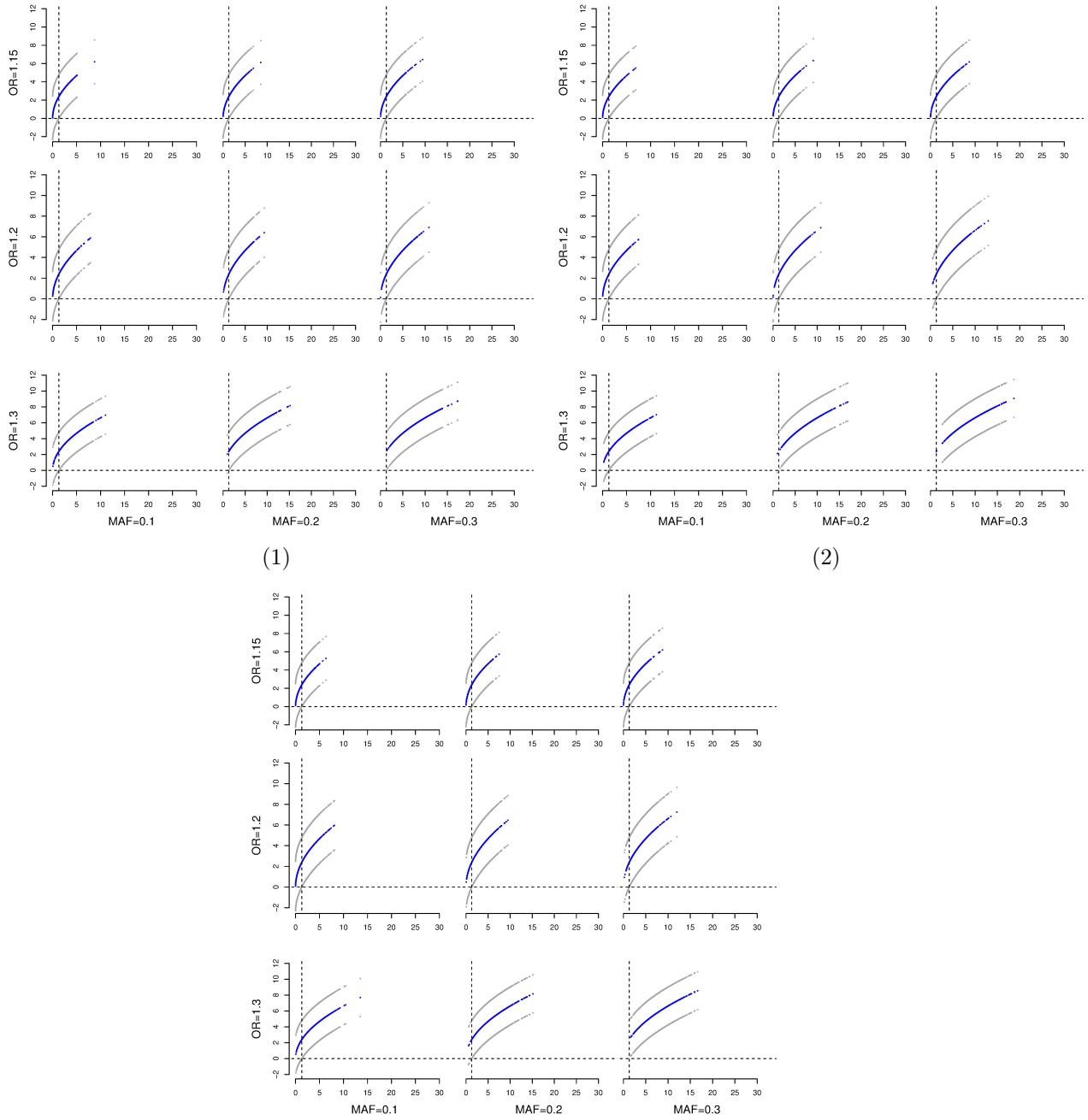
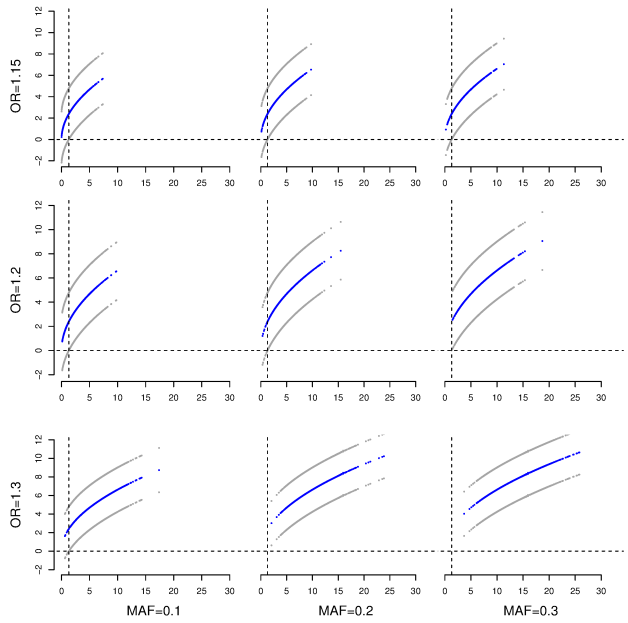
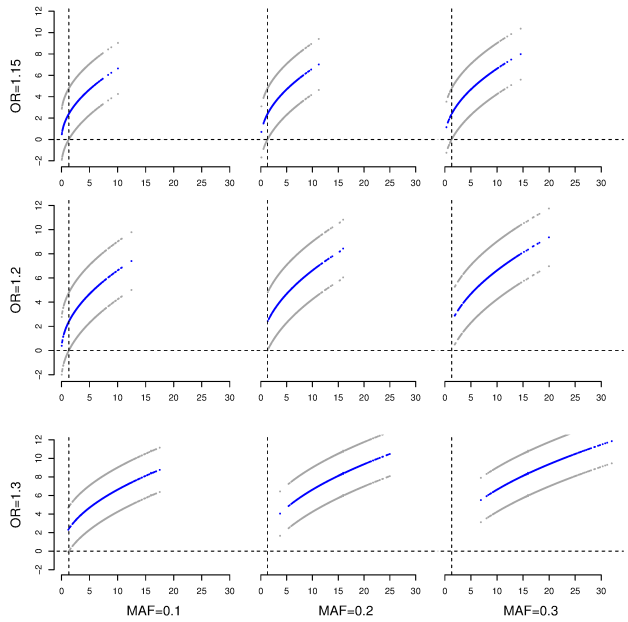


Figure 3.20.: **Settings 1-3. Agonistic effect.** Simulation-based comparison of $-log_{10}(P)$ (x axis) vs. 95% CI (y axis) for detecting agonistic effects with $P < 0.05$. Blue dots depict $T_{\max}$ and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (1) unequal baseline risks for the two diseases of 0.1 and 0.05, respectively, equally distributed controls; (2) equal baseline risk of 0.1 for both diseases, equally distributed controls; (3) unequal baseline risks for the two diseases of 0.1 and 0.05, respectively, proportionally distributed controls.

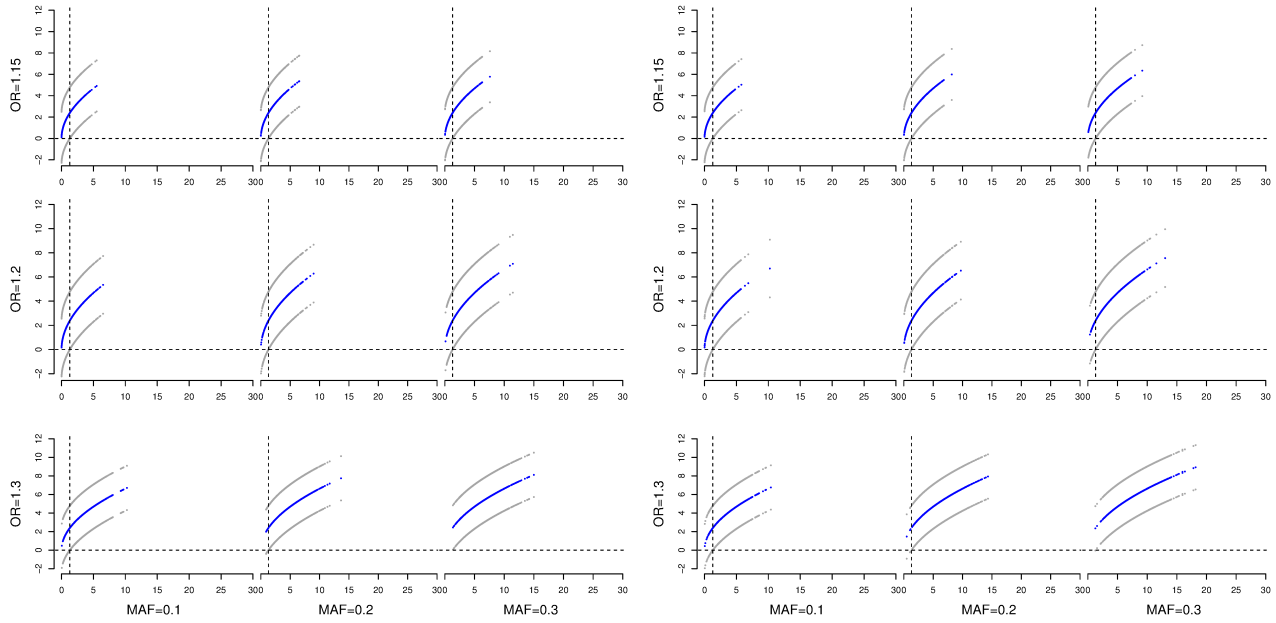


(4)

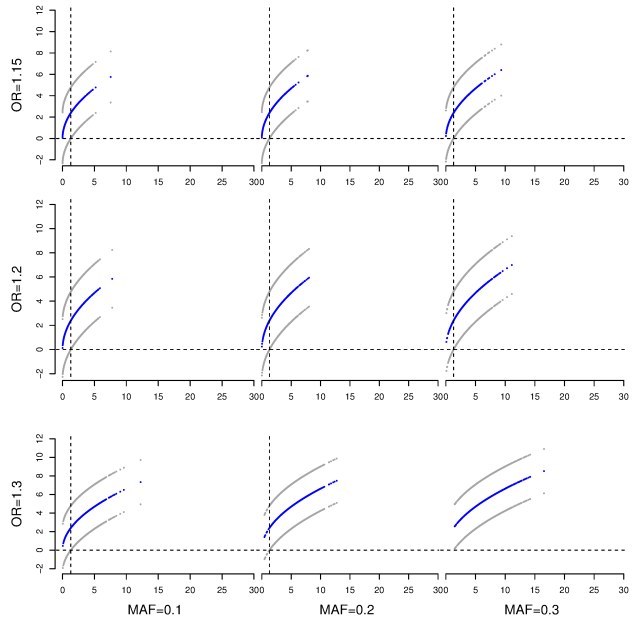


(5)

Figure 3.21.: **Settings 4-5. Agonistic effect.** Simulation-based comparison of $-log_{10}(P)$ (x axis) vs. 95% CI (y axis) for detecting agonistic effects with $P < 0.05$. Blue dots depict $T_{\max}$ and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (4) unequal baseline risks for the two diseases of 0.1 and 0.05, respectively, case-control ratio=1:2; (5) equal baseline risk of 0.1 for both diseases, case-control ratio=1:2.

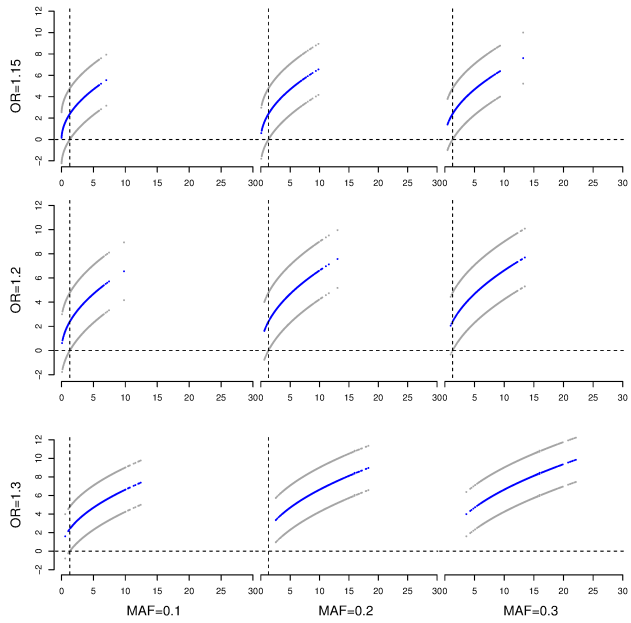


(1) (2)

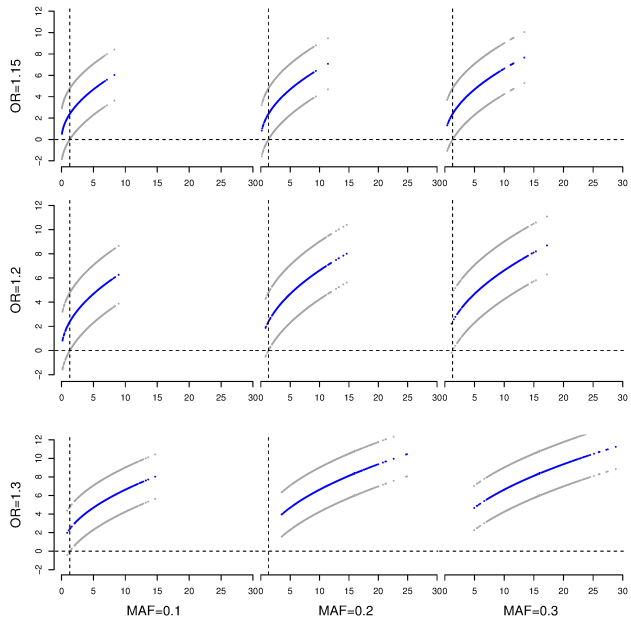


(3)

Figure 3.22.: **Settings 1-3. Antagonistic effect.** Simulation-based comparison of $-log_{10}(P)$ (x axis) vs. 95% CI (y axis) for detecting antagonistic effects with $P < 0.05$. Blue dots depict $T_{\max}$ and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (1) unequal baseline risks for the two diseases of 0.1 and 0.05, respectively, equally distributed controls; (2) equal baseline risk of 0.1 for both diseases, equally distributed controls; (3) unequal baseline risks for the two diseases of 0.1 and 0.05, respectively, proportionally distributed controls.



(4)



(5)

Figure 3.23.: **Settings 4-5. Antagonistic effect.** Simulation-based comparison of $-log_{10}(P)$ (x axis) vs. 95%CI (y axis) for detecting antagonistic effects with $P < 0.05$. Blue dots depict $T_{\max}$ and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (4) unequal baseline risks for the two diseases of 0.1 and 0.05, respectively, case-control ratio=1:2; (5) equal baseline risk of 0.1 for both diseases, case-control ratio=1:2.

3.2.2. Confidence Intervals for the Effect of Interest

The confidence interval derived by eq. (2.31) for the disease specific effect $\hat{\theta}_1$ or $\hat{\theta}_2$, if one of them yields the maximum statistic, is symmetric but wider than the one usually calculated by the normal approximation. This reflects the greater variability by searching for the maximum test statistic and leads to a more conservative confidence bound. It has been shown above that the proportion of 95% confidence interval estimates for the test statistic $T_{\max} > 0$ corresponds with the power of the hypothesis test on a 5% significance level. The duality of the P-value with the 95% confidence interval is also observed for the disease specific effect estimate (Figures 3.24, 3.25).
When studying binary traits, the disease specific effect estimate and its confidence interval can be transformed by $exp(\hat{\theta})$ and interpreted as a confidence interval for an odds ratio, which is more conservative than the one obtained by the normal approximation.
On the contrary, the interpretation of the confidence interval for agonistic and antagonistic effects is not straightforward. As can be seen in eq. (2.33), the numerator for the agonistic or antagonistic effect is a weighted sum or difference, whereas the weights are attenuated towards one by the re-definition in eq. (2.32). Thus, the absolute sum or difference cannot be extracted, except for the special case when both variances are equal. However, this estimator taken from the numerator of eq. (2.32) reflects almost the sum or the difference. This is supported by the results shown in Figures 3.26-3.29 revealing that the magnitude of those weighted effect estimates is about twice the size of those obtained for the disease specific effect and thus comparable.
For the agonistic or antagonistic effect one would be rather interested to see the strength of the SNP effect simultaneously on each of the two diseases with corresponding confidence intervals. This can be done using the multinomial regression model as described in Section 2.5 when individual data are available.

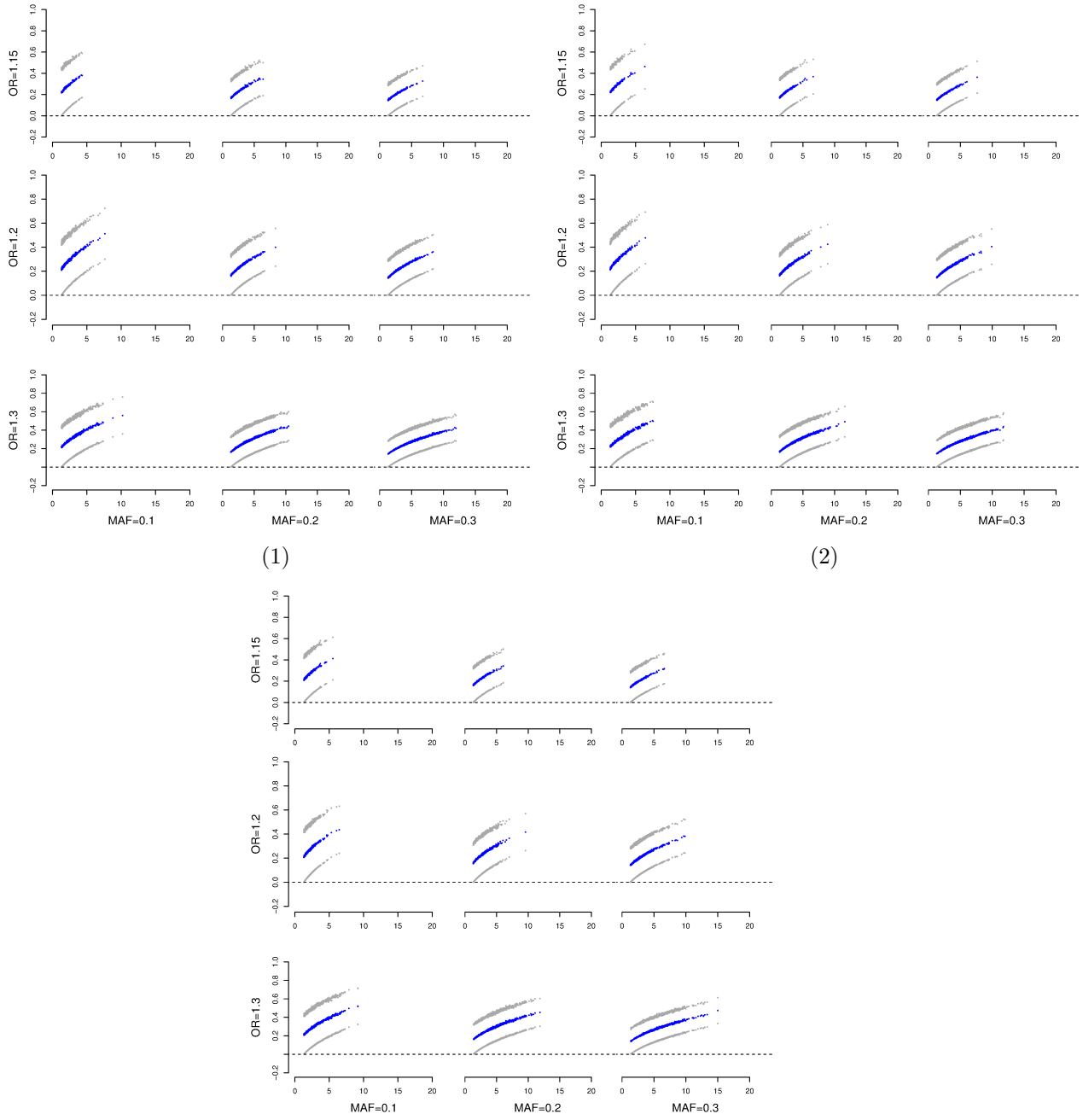
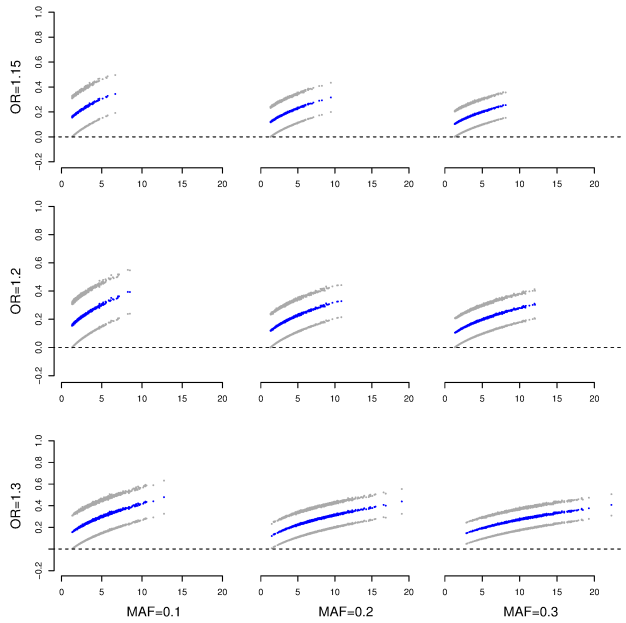
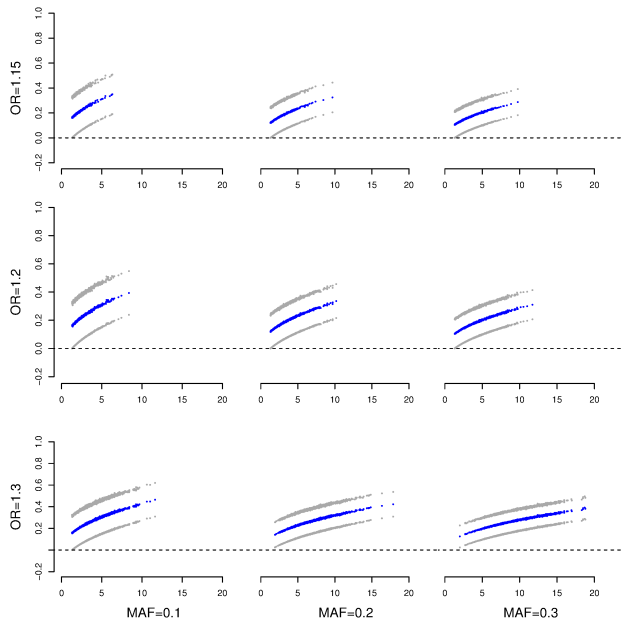


Figure 3.24.: **Settings 1-3. Disease specific effect.** $-log_{10}(P)$ (x axis) vs. 95% CI of the effect of interest (y axis) of simulated data with $P < 0.05$. Blue dots depict the effect of interest and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (1) unequal baseline risk of 0.1 and 0.05, equally distributed controls; (2) equal baseline risk of 0.1, equally distributed controls; (3) unequal baseline risk of 0.1 and 0.05, proportionally distributed controls.



(4)



(5)

Figure 3.25.: **Settings 4-5. Disease specific effect.** $-log_{10}(P)$ (x axis) vs. 95% CI of the effect of interest (y axis) of simulated data with $P < 0.05$. Blue dots depict the effect of interest and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (4) unequal baseline risk of 0.1 and 0.05, case-control ratio=1:2; (5) equal baseline risk of 0.1, case-control ratio=1:2.

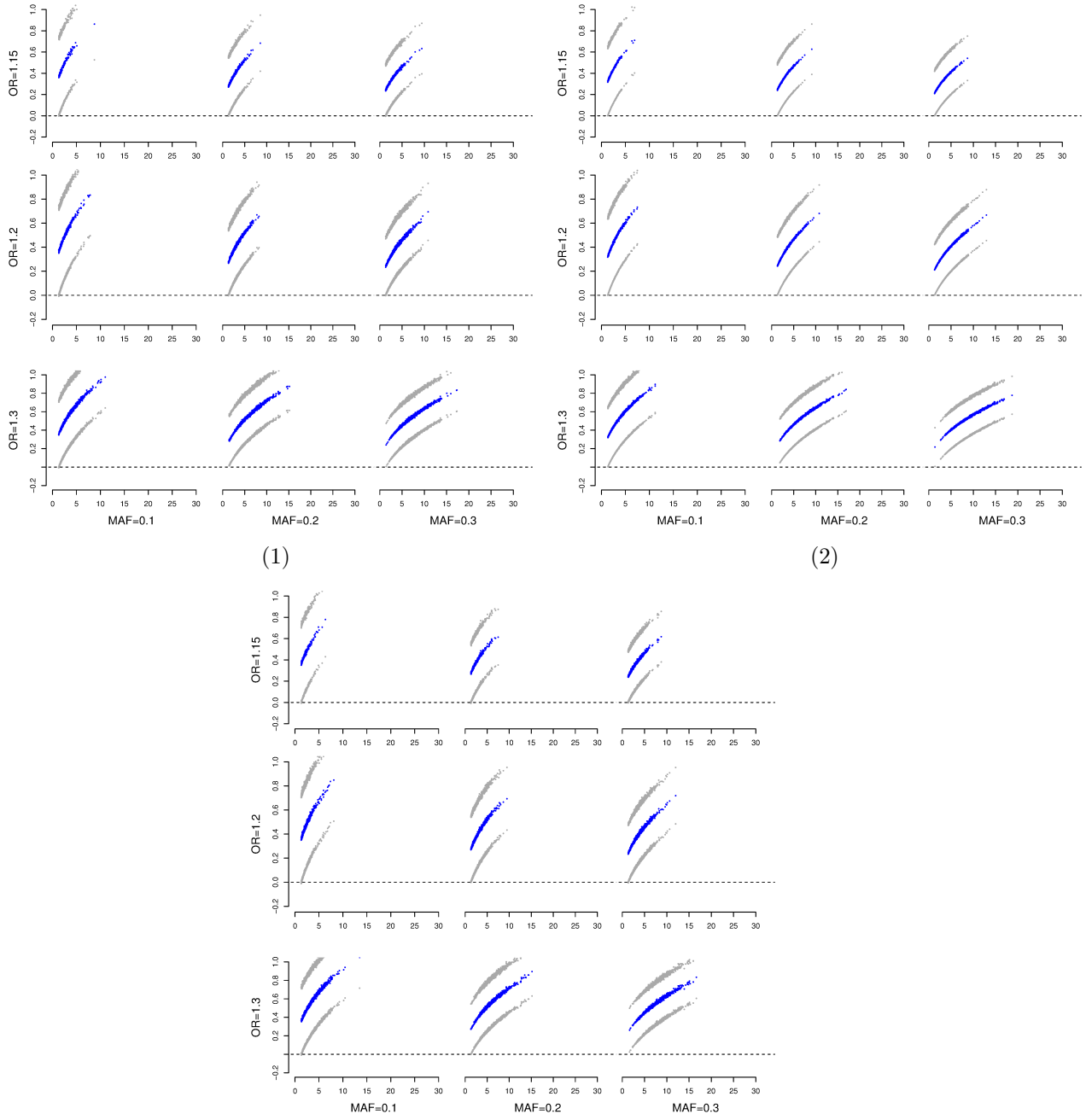
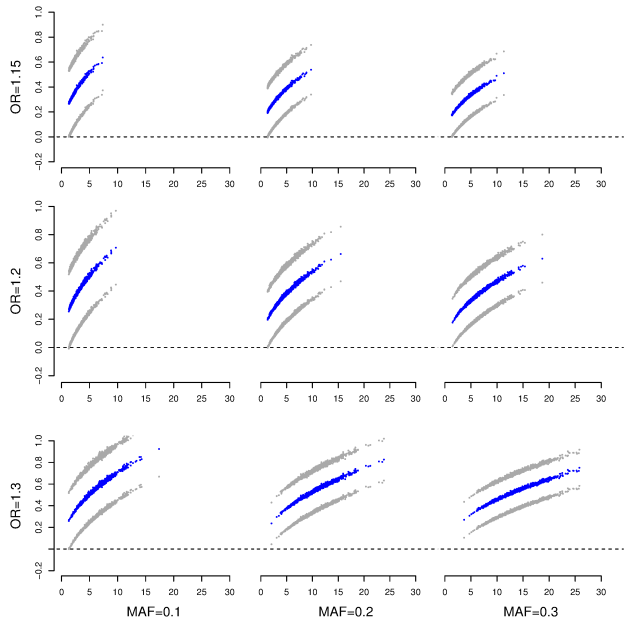
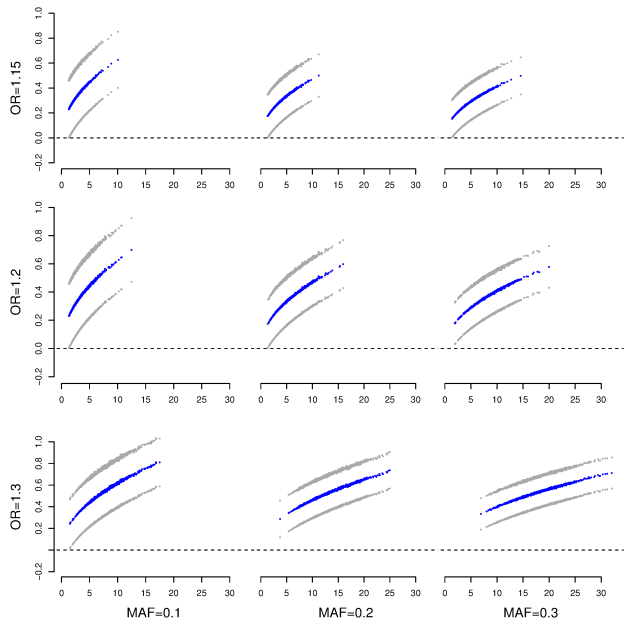


Figure 3.26.: **Settings 1-3. Agonistic effect.** $-log_{10}(P)$ (x axis) vs. 95% CI of the effect of interest (y axis) of simulated data with $P < 0.05$. Blue dots depict the effect of interest and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (1) unequal baseline risk of 0.1 and 0.05, equally distributed controls; (2) equal baseline risk of 0.1, equally distributed controls; (3) unequal baseline risk of 0.1 and 0.05, proportionally distributed controls.



(4)



(5)

Figure 3.27.: **Settings 4-5. Agonistic effect.** $-log_{10}(P)$ (x axis) vs. 95% CI of the effect of interest (y axis) of simulated data with $P < 0.05$. Blue dots depict the effect of interest and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (4) unequal baseline risk of 0.1 and 0.05, case-control ratio=1:2; (5) equal baseline risk of 0.1, case-control ratio=1:2.

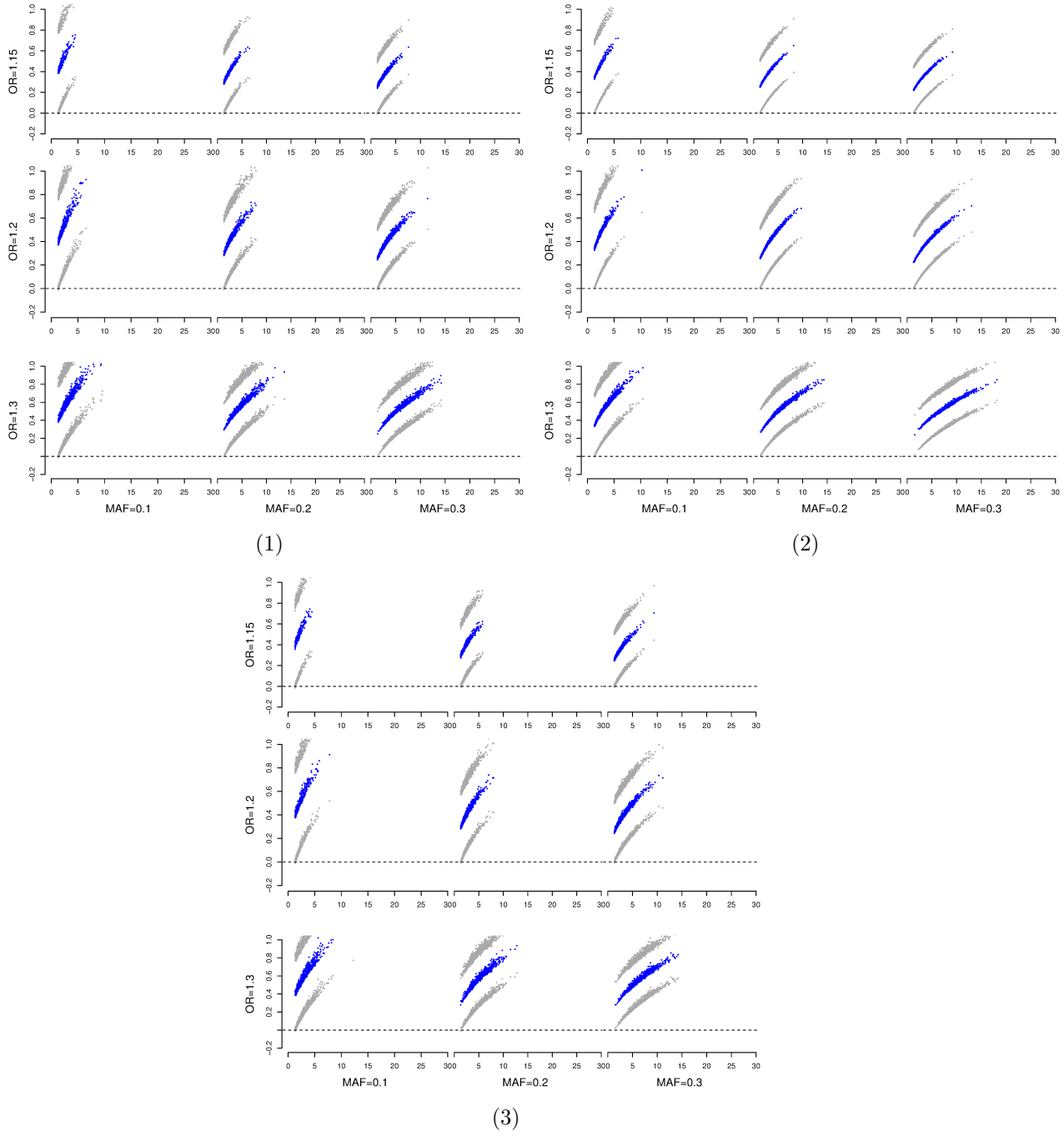
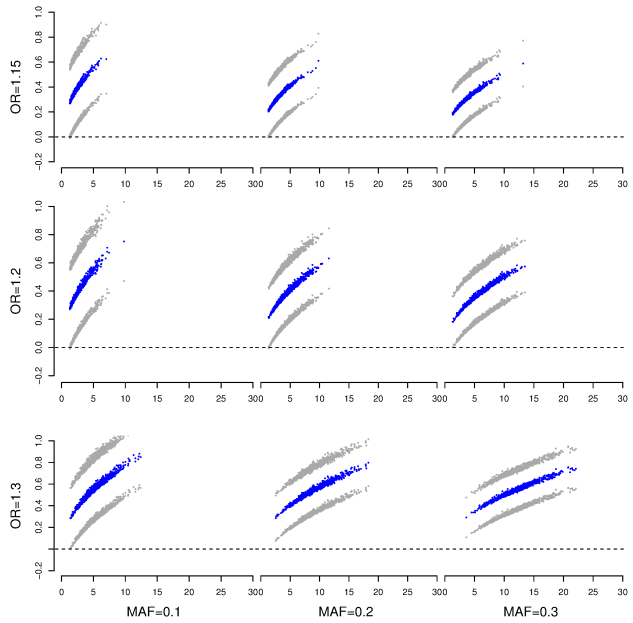
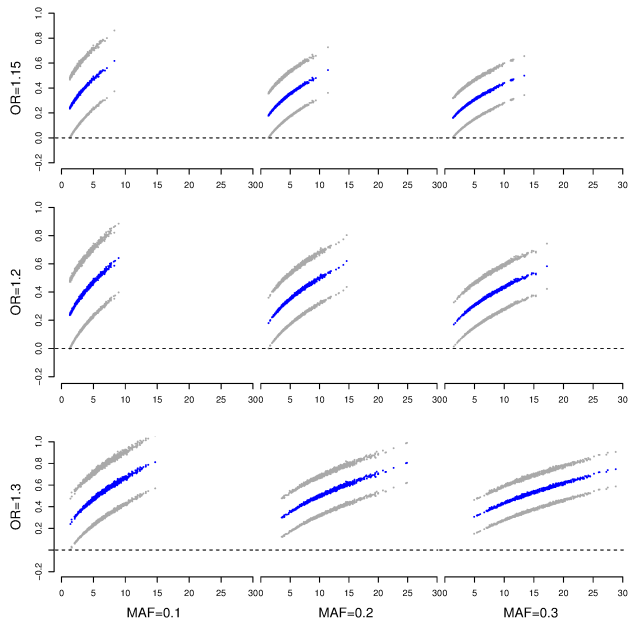


Figure 3.28.: **Settings 1-3. Antagonistic effect.** $-log_{10}(P)$ (x axis) vs. 95% CI of the effect of interest (y axis) of simulated data with $P < 0.05$. Blue dots depict the effect of interest and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (1) unequal baseline risk of 0.1 and 0.05, equally distributed controls; (2) equal baseline risk of 0.1, equally distributed controls; (3) unequal baseline risk of 0.1 and 0.05, proportionally distributed controls.



(4)



(5)

Figure 3.29.: **Settings 4-5. Antagonistic effect.** $-log_{10}(P)$ (x axis) vs. 95% CI of the effect of interest (y axis) of simulated data with $P < 0.05$. Blue dots depict the effect of interest and grey dots the upper and lower bound of the respective 95% CI. Rows indicate the specified OR and columns the specified MAF. (4) unequal baseline risk of 0.1 and 0.05, case-control ratio=1:2; (5) equal baseline risk of 0.1, case-control ratio=1:2.

3.3. CCMA for Comparing and Contrasting AE and Psoriasis

After imputation three GWA studies for AE, totaling in 2,079 cases vs. 3,867 controls, and three GWA studies for psoriasis, totaling in 4,212 cases vs. 8,032 controls, have been meta-analyzed in the first place. For comparing and contrasting AE and psoriasis, we then used the CCMA method and identified 24,187 agonistic, antagonistic or disease specific SNVs with $T_{max} > 4.68$ (Figure 3.30). This critical value corresponds to a suggestive significance threshold of $P < 10^{-5}$, which was chosen to reduce the probability of false-negatives (see Figure 2.3).
The MHC region is spared for a separate analysis because of its unique and complex variability and patterns of strong linkage disequilibrium (LD). After exclusion of the MHC we identified 2,260 agonistic, antagonistic or disease-specific SNVs. The 2,260 SNVs were condensed to 142 distinct loci after an LD-based clumping procedure[115] using the parameters distance $\leq$ 250kb and $r^2 \geq 0.5$, i.e. all SNVs within a defined region with LD-values above the defined threshold are assigned to the SNV with smallest P-value (lead SNV) representing the association with the disease (see Supplementary Table A.2). The filtered SNVs showing a suggestive association by the CCMA method are carried forward for multinomial regression modeling (MNM) to estimate effect sizes for both diseases simultaneously.

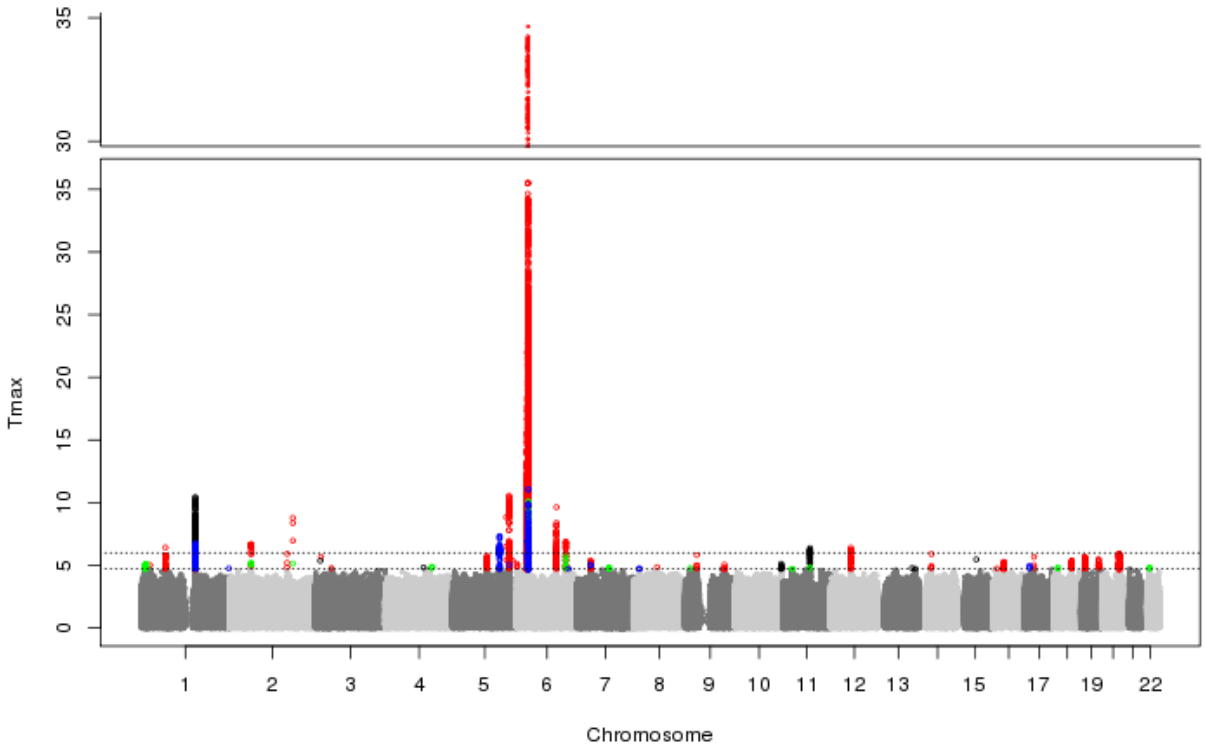


Figure 3.30.: Manhattan plot for psoriasis and atopic eczema using CCMA. The effect categorization is color coded: **black**=atopic eczema, **red**=psoriasis, **blue**=antagonistic, **green**=agonistic effects and **grey**=depicts none significant signals.

3.4. Comparison with Multinomial Regression Analysis

For evaluating the ability of the CCMA method to correctly categorize SNVs into single disease or pleiotropic effects, we compare the effect categorization with the categorization derived by the multinomial regression model as a gold standard as described in Section 2.7. To this end, individual genotype data of the 24,032 SNVs with suggestive association identified by SBMA of all 6 studies are extracted, and a complete data set of 2,079 atopic eczema cases, 4,212 psoriasis cases and 11,899 controls is created.
The 24,032 selected SNVs are carried forward to multinomial regression modeling adjusted for age, sex and the first four genome-wide principal components (see Table 2.2).
With the use of (2.12) and (2.37) we derive the effect categorizations from CCMA and MNM, respectively, and compare the concordance in a contingency table (see Section 2.7). Using formula (2.38) we calculate a concordance rate of both methods of 85.5% and 88.2% for SNVs with suggestive ($P_{\text{SBMA}} < 10^{-5}$) and genome-wide significant association ($P_{\text{SBMA}} < 10^{-8}$) in SBMA, respectively (Table 3.11).

	MNM							
CCMA	AE		Agonistic		Antagonistic		Psoriasis	
	10^{-5}	10^{-8}	10^{-5}	10^{-8}	10^{-5}	10^{-8}	10^{-5}	10^{-8}
AE	**270**	**234**	18	0	39	18	3	0
Agonistic	33	12	**123**	**43**	14	5	284	215
Antagonistic	11	1	9	0	**751**	**439**	306	213
Psoriasis	128	54	632	282	2014	1344	**19397**	**15350**

Table 3.11.: Concordance of effect categorizations between CCMA and MNM of all SNVs with suggestive ($P_{\text{SBMA}} < 10^{-5}$) and genome-wide significant association ($P_{\text{SBMA}} < 10^{-8}$) in SBMA. Concordant categorizations are indicated by bold numbers. The resulting concordance rates are 85.5% and 88.2%, respectively.

When excluding SNVs from the complex MHC region, the concordance rate of both methods increases to 93.6% and 96.6% for suggestive and genome-wide significant association in SBMA, respectively (Table 3.12, Figure 3.31).

	MNM							
CCMA	AE		Agonistic		Antagonistic		Psoriasis	
	10^{-5}	10^{-8}	10^{-5}	10^{-8}	10^{-5}	10^{-8}	10^{-5}	10^{-8}
AE	**268**	**234**	16	0	28	18	3	0
Agonistic	18	6	**30**	**0**	0	0	23	1
Antagonistic	5	0	1	0	**403**	**236**	6	0
Psoriasis	1	0	11	0	23	2	**1283**	**288**

Table 3.12.: Concordance of effect categorizations between CCMA and MNM of all SNVs with suggestive ($P_{\text{SBMA}} < 10^{-5}$) and genome-wide significant association ($P_{\text{SBMA}} < 10^{-8}$) in SBMA excluding the MHC region. Concordant categorizations are indicated by bold numbers. The resulting concordance rates are 93.6% and 96.6%, respectively.

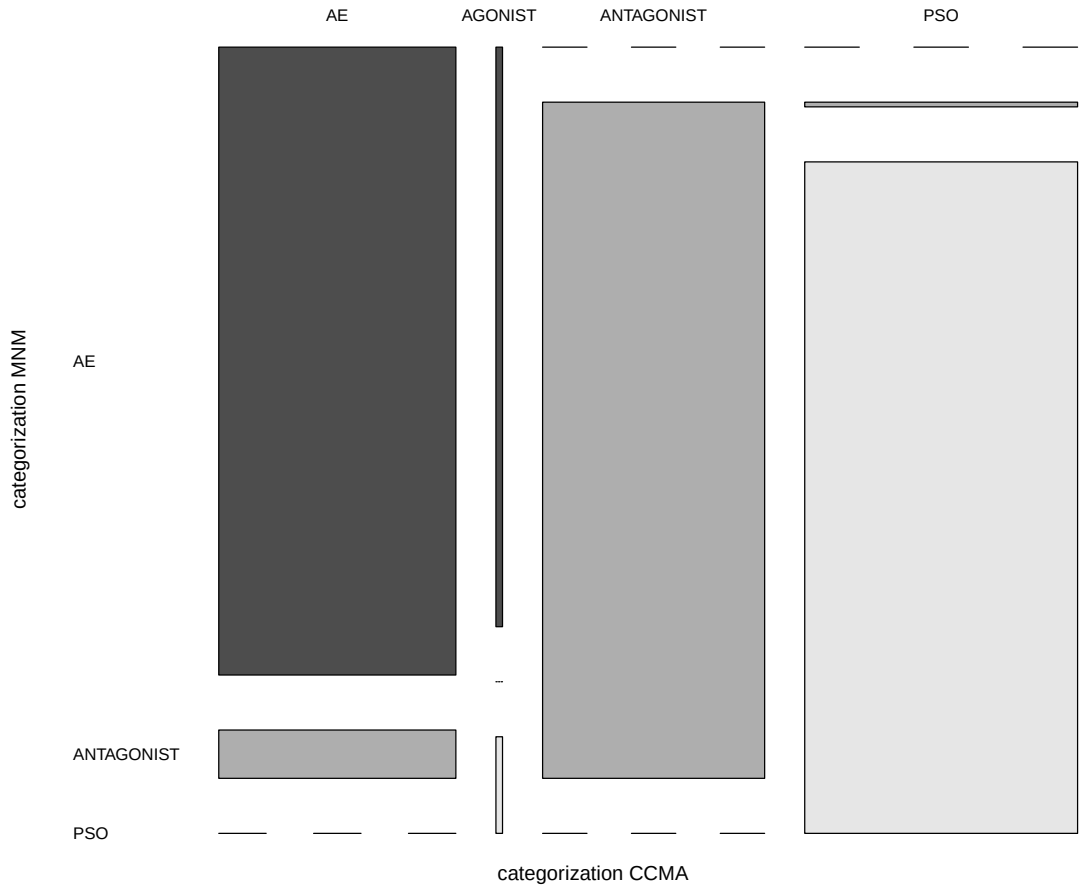


Figure 3.31.: Mosaic plot of concordance between CCMA and MNM effect categorizations of SNVs with genome-wide significant association detected by SBMA ($P_{\text{SBMA}} < 10^{-8}$) excluding the MHC region.

3.5. Comparison with the COMBINED & OVERLAP Method

The COMBINED & OVERLAP procedure has been applied to the meta analysis results on atopic eczema and psoriasis. Of the 36 and 21 previously established risk SNVs for psoriasis[2] and atopic eczema,[3,98,116] respectively, only 24 and 11 SNVs are available in our dataset and hence meta-analysis results only of those could be obtained.
Using the proposed thresholds (see section 2.6), the OVERLAP reveals only one of the psoriasis risk SNVs and only three of the AE risk SNVs to be associated with the other trait. Therefore the known risk loci are defined in a broader sense including all SNVs ± 500kb around the known risk SNVs meeting a genome-wide significance threshold of $P < 10^{-8}$ in the respective disease-specific meta-analysis. This definition results in 32 AE and 1,598 psoriasis risk SNVs for the vice versa lookup. Applying the OVERLAP approach to this enlarged SNV sets, 732 psoriasis SNVs ($P < 10^{-8}$), which all map to the MHC, show significance for atopic eczema ($P < 0.01$), and 7 AE SNVs ($P < 10^{-8}$) are also associated with psoriasis ($P < 0.01$) and map to the EDC region on 1q21.3, *RAD50/IL13* on 5q31.1 and the MHC on 6p21.33.
Conducting the COMBINED approach for the same and opposite effect direction as described in Section 2.6 ($P_{\text{atopic eczema}} < 0.05$ & $P_{\text{psoriasis}} < 0.05$ & $P_{\text{COMBINED}} < 10^{-4}$) reveals 2,112 SNVs with effects in the same direction and 4,285 SNVs with effects in the opposite direction. However, the SNVs are spread across all chromosomes. Of note, all SNVs detected by the OVERLAP approach are also detected by the COMBINED approach. Hence, only the COMBINED approach will be compared with CCMA and MNM.
Applying a more stringent filter using a suggestive threshold of $P_{\text{COMBINED}} < 10^{-5}$ reveals 1,590 SNVs for the same effect direction and 3,606 SNVs for opposite effect direction (Table 3.13). Again, all SNVs detected by the OVERLAP approach are included in the set of SNVs detected by the COMBINED approach.
Comparison of effect categorization of the COMBINED method with CCMA and MNM was again based on the set of 24,032 SNVs with suggestive association detected by SBMA. Of these, 19,179 SNVs are not retrieved by the COMBINED method, resulting in a set of 4,853 intersecting SNVs that are subsequently used for the comparison of effect categorization.
The concordance rate of COMBINED and CCMA is 83.3% while the concordance rate with MNM, the gold standard, is only 70.9% (Table 3.14). In contrast, the concordance rate of CCMA and MNM for the same set of SNVs is 80.5% (Figure 3.32).
When discarding the MHC region, the set of overlapping SNVs reduces to 522 SNVs but reveals higher concordance rates between all three methods. The concordance rate of the COMBINED analysis with CCMA and MNM is 93.3% and 85.4%, respectively (Table 3.15). The concordance rate of CCMA with MNM is 89.5% (data not shown).

Chromosome	Effect direction		Overall
	same	opposite	
1	45	310	355
2	8	-	8
4	10	-	10
5	5	119	124
6	1454	3161	4615
7	14	5	19
8	-	6	6
9	3	-	3
11	33	1	34
12	2	-	2
13	-	1	1
17	-	3	3
18	14	-	14
20	2	-	2
Sum	1590	3606	5196

Table 3.13.: Number of pleiotropic SNVs identified for AE and psoriasis using the COMBINED method and restricting to SNVs meeting the threshold of suggestive association ($P < 10^{-5}$).

COMBINED	AE		Agonistic		Antagonistic		Psoriasis	
	CCMA	**MNM**	**CCMA**	**MNM**	**CCMA**	**MNM**	**CCMA**	**MNM**
AE	**24**	**19**	6	0	6	16	0	1
Agonistic	0	28	**422**	**150**	0	24	156	376
Antagonistic	0	7	0	11	**1041**	**840**	609	792
Psoriasis	0	6	13	24	17	128	**2559**	**2431**

Table 3.14.: Concordance of effect categorization between the COMBINED method with CCMA and MNM based on the common set of SNVs with suggestive association ($P < 10^{-5}$) detected by COMBINED and SBMA. Concordant categorizations are indicated by bold numbers. The resulting concordance rates of the COMBINED method with CCMA and MNM are 83.3% and 70.9%, respectively.

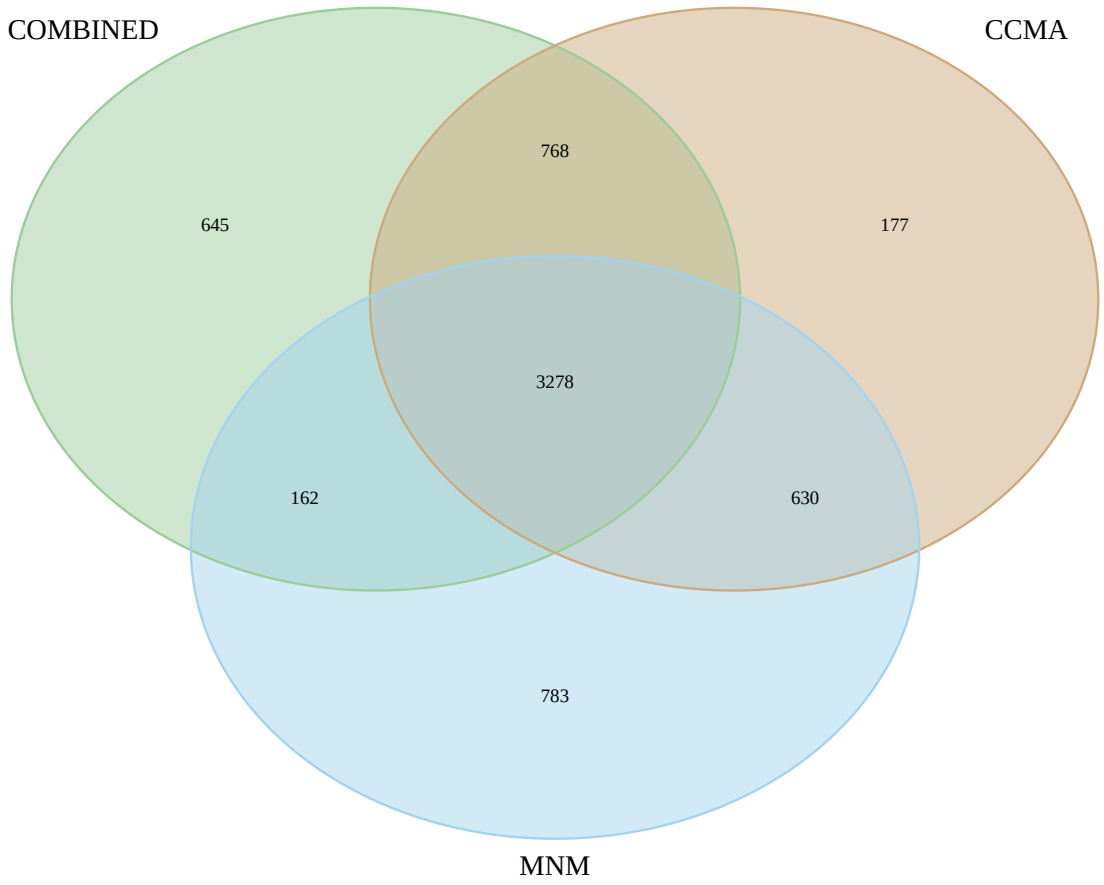


Figure 3.32.: Venn diagram showing the concordance of COMBINED, CCMA and MNM effect categorization for the common set of 4,853 SNVs with suggestive association ($P < 10^{-5}$) detected by COMBINED and SBMA.

COMBINED	AE		Agonistic		Antagonistic		Psoriasis	
	CCMA	**MNM**	**CCMA**	**MNM**	**CCMA**	**MNM**	**CCMA**	**MNM**
AE	**24**	**19**	6	0	6	16	0	1
Agonistic	0	12	**59**	**29**	0	0	20	38
Antagonistic	0	0	0	1	**401**	**395**	3	8
Psoriasis	0	0	0	0	0	0	**3**	**3**

Table 3.15.: Concordance of effect categorization between the COMBINED method with CCMA and MNM based on the common set of SNVs with suggestive association ($P < 10^{-5}$) detected by COMBINED and SBMA excluding the MHC region. Concordant categorizations are indicated by bold numbers. The resulting concordance rates of the COMBINED method with CCMA and MNM are 93.3% and 85.4%, respectively.

Investigating the set of 4,124 SNVs with genome-wide association in the SBMA approach ($P_{\text{SBMA}} < 10^{-8}$) reveals concordance rates of 84.1% and 73.4% between COMBINED and CCMA and MNM, respectively (Table 3.16). The concordance rate between CCMA and MNM for the same set of SNVs is 82.6% (data not shown).

COMBINED	AE		Agonistic		Antagonistic		Psoriasis	
	CCMA	**MNM**	**CCMA**	**MNM**	**CCMA**	**MNM**	**CCMA**	**MNM**
AE	**17**	**14**	6	0	1	10	0	0
Agonistic	0	7	**258**	**65**	0	8	87	265
Antagonistic	0	2	0	2	**640**	**520**	536	652
Psoriasis	0	6	11	24	12	122	**2556**	**2427**

Table 3.16.: Concordance of effect categorization between the COMBINED method with CCMA and MNM based on the set of SNVs with genome-wide association ($P < 10^{-8}$) detected by SBMA. Concordant categorizations are indicated by bold numbers. The resulting concordance rates of the COMBINED method with CCMA and MNM are 84.1% and 73.4%, respectively.

3.6. Comparative Analysis of AE and Psoriasis Gives Insight into Opposing Genetic Mechanisms

With the CCMA 24,187 SNVs with suggestive association ($T_{max} > 4.68$) have been identified, which can be condensed to 142 distinct loci by LD pruning after excluding the MHC (see Section 3.3). No agonistic loci with genome-wide association ($T_{\max}$) are found, but five loci which are co-associated with both diseases and have antagonistic effects (see Supplementary Table A.2).
The GWAS data have subsequently been analyzed for the identified loci using the MNM. For validation, these regions of interest have also been investigated by MNM using data, obtained with the Immunochip, which is a customized array designed for fine mapping established GWAS loci of immune-mediated diseases. The 2,425 AE patients, 3,580 psoriasis patients and 9,061 controls were obtained from previous studies[3,2] including data on a subset of case and control individuals also analyzed by GWAS.

3.6.1. Identification of New Opposing Loci for AE and Psoriasis

Excluding the MHC, four loci show co-association with AE and psoriasis and are located in the chromosomal regions 1q21.3, 2q31.2, 5q31.1 and 5q33.1. The epidermal differentiation complex (EDC) region (1q21.3) and the cytokine cluster (5q31.1) are established risk loci for both diseases and are characterized by complex LD structure. The EDC harbors the *FLG* gene, which is "the strongest known genetic risk factor for AE and encodes a key epidermal structural protein".[117] Therefore these loci are interrogated by stepwise conditional analysis using the multinomial regression model to dissect these loci into disease specific and antagonistic effects.
Rs62176107 on 2q31.2 demonstrates an antagonistic effect ($P_{\text{MNM}} = 1.08 \times 10^{-34}$) and is within exon 6 of *PRKRA*. *PRKRA* encodes an activator of the protein kinase R enzyme in response to extracellular stress[118] and has been shown to be an essential factor of the RIG-I-mediated antiviral response.[119]
The lead SNV rs17728338 at 5q33.1 with antagonistic effects on AE and psoriasis ($P_{\text{MNM}} = 3.96 \times 10^{-38}$) is located in a 25-kb LD-block containing both *ANXA6* and *TNIP1*. *TNIP1* has previously been reported as a psoriasis locus in European and Chinese populations, but not with AE. *TNIP1* plays a role in TNF signaling and the regulation of the transcription factor NFκB[120] and shows increased expression in lesional skin of psoriasis and AE patients compared to controls.[106] In contrast, *ANXA6*, which encodes a calcium-dependent membrane and phospholipid binding protein, shows a 1.3-fold increased expression in atopic compared to normal skin, whereas in psoriatic skin it is 0.7-fold decreased compared to normal skin.[106]

Conditional Analysis of 1q21.3 and 5q31.1

LD analysis reveals seven blocks with disease-specific or antagonistic signals within 1q21.3 (Figures 3.33 and 3.34). Conditional analysis on the four most prevalent *FLG* null mutations, which are strong risk factors for AE, and rs1581803 tagging the *LCE3B-LCE3C*

deletion, which is strongly associated with psoriasis, identifies a locus with antagonistic effects mapping to *RPTN/HRNR/*FLG-AS1. The G allele of the lead SNV rs12130219 shows decreased risk for AE ($OR_{AE\ cond} = 0.81$, $P_{AE\ cond} = 0.0018$) and increased risk for psoriasis ($OR_{PSO\ cond} = 1.12$, $P_{PSO\ cond} = 3.68 \times 10^{-4}$) (Table 3.17). Each of the proteins filaggrin, repetin and hornerin contribute to the cornified cell envelope, which is a functional component of the epidermal barrier. *FLG* and *HRNR* expression has been shown to be up-regulated in AE[121,122] whilst it may be dys- or down-regulated in psoriasis.[123,124] The proximity of FLG-AS1 to *FLG* and *HRNR* suggests a role in coordinating keratinocyte terminal differentiation, however its function is currently unknown. Increased expression of FLG-AS1 is observed in lesional psoriatic skin and decreased expression in lesional AE skin compared to non-lesional skin.[106]

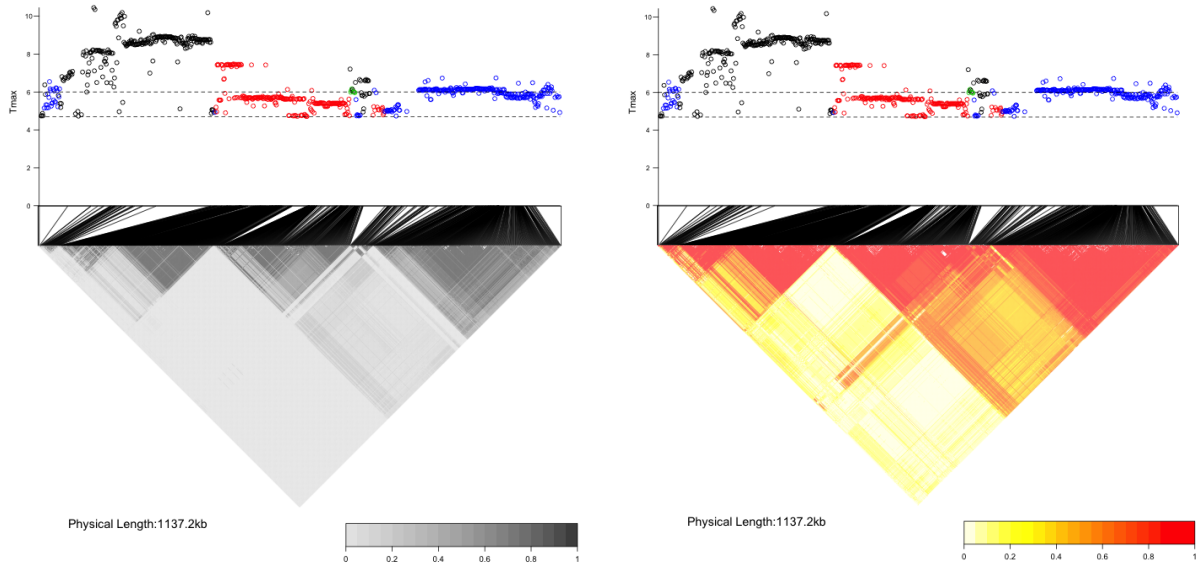


Figure 3.33.: LD structure of all SNVs in the EDC region on chromosome 1q21.3 meeting the threshold of suggestive association $T_{max} > 4.68$. The left figure shows r^2 and the right figure shows D' values.

At 5q31.1 an antagonistic effect for the locus with the strongest association is identified. The lead SNV rs6596086 maps to *RAD50* and its C allele is associated with an increased risk to AE and a decreased risk to psoriasis ($OR_{AE\ full} = 1.17$, $OR_{PSO\ full} = 0.88$, $P_{overall} = 6.3 \times 10^{-7}$). Stepwise conditional analysis was carried out and after the final selection step a full model was calculated. The analysis reveals three additional independent loci exclusively associated with AE: *IL13* (rs848, $OR_{AE\ full} = 1.12$, $P_{AE\ full} = 0.0197$), *KIF3A* (rs2299009, $OR_{AE\ full} = 1.16$, $P_{AE\ full} = 2.0 \times 10^{-4}$) and *SLC22A4 / C5orf56* (rs74458173, $OR_{AE\ full} = 1.57$, $P_{AE\ full} = 5.7 \times 10^{-5}$) (Figure 3.35, Table 3.17).

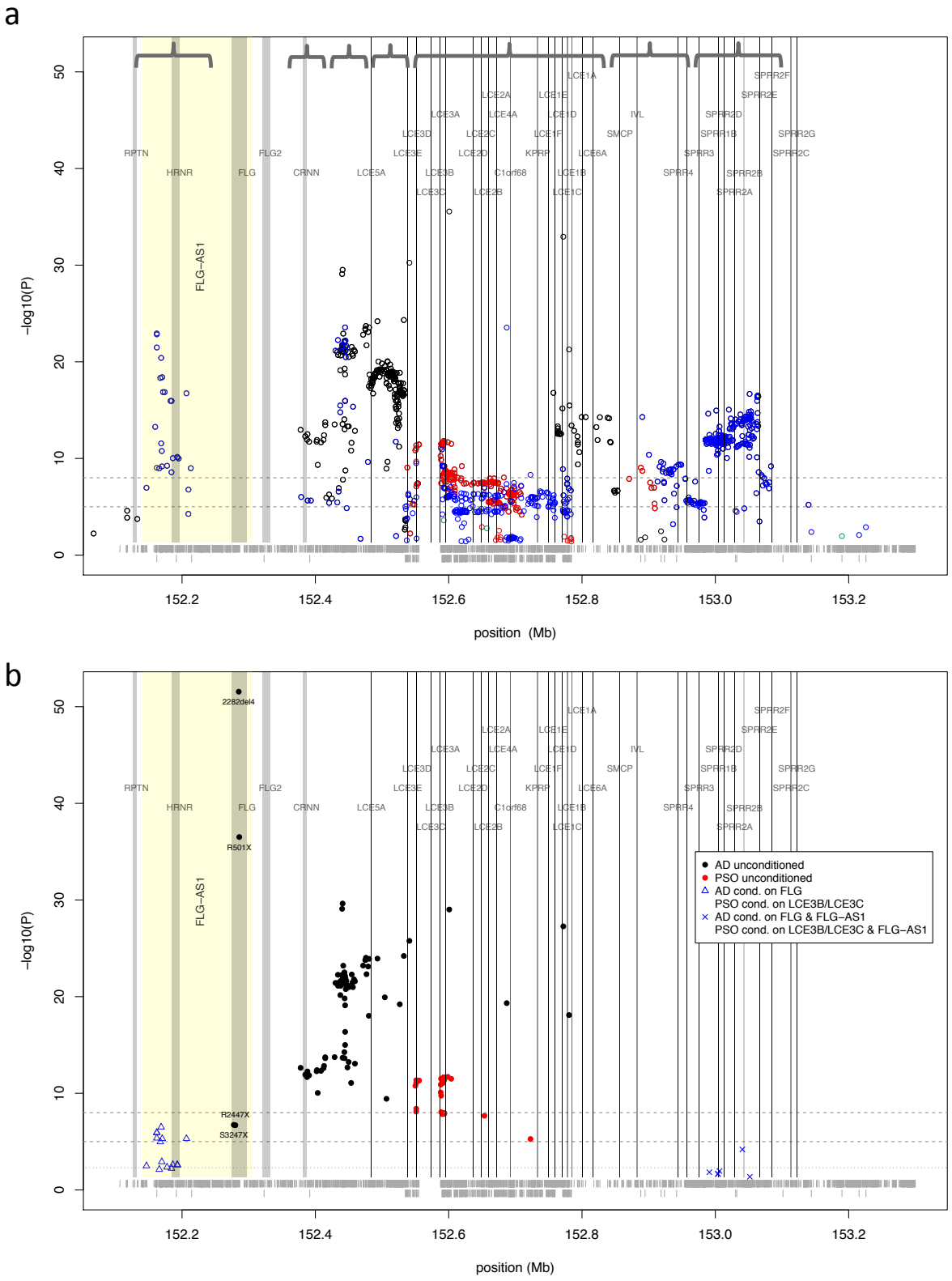


Figure 3.34.: Regional association within the epidermal differentiation complex at 1q21.3. (a) Multinomial regression model using GWAS and ImmunoChip data. Seven blocks of linkage disequilibrium are indicated by curly brackets; **black** circles indicate AE-specific, **red** psoriasis-specific and **blue** antagonistic SNV effects. (b) Conditional regional association plot of stepwise logistic regression. The different colored symbols indicate association results after each step of analysis, as follows: Unconditioned results are shown by **black** and **red** dots to indicate association with AE and psoriasis, respectively; **blue** triangles and **blue** crosses represent results after conditioning on the known disease-associated variants, *FLG* in AE and *LCE3B-LCE3C* deletion in psoriasis; SNVs indicated by the same symbol are in LD. For both plots, vertical lines are drawn to mark the positions of known genes (Genome Reference Consortium GRCh37) and transcripts, and horizontal dotted lines indicate significance thresholds of $P = 0.005$, 10^{-5} and 10^{-8}. The horizontal bands at the bottom indicate the coverage of the region by GWAS SNVs (upper row) and ImmunoChip SNVs (lower row).

Chromsome 1q21.3

Data source	Effect	SNV	Position	Allele	Candidate Gene	Unconditional analysis					Conditional analysis[a]				
						AE		Psoriasis		$P_{overall}$	AE		Psoriassis		$P_{overall}$[d]
						OR (95% CI)	P	OR (95% CI)	P		OR (95% CI)	P	OR (95% CI)	P	
GWAS	antagonistic	rs12130219	152162106	G/A	FLG-AS1 *RPTN/HRNR*	0.66 (0.60-0.73)	$1.1x10^{-16}$	1.15 (1.09-1.22)	$4.0x10^{-6}$	$1.2x10^{-23}$	0.81 (0.71-0.93)	0.0018	1.12 (1.05-1.19)	$3.7x10^{-4}$	$2.4x10^{-6}$
GWAS	AE	rs12144049	152440910	C/T	*FLG*	1.53 (1.42-1.64)	$2.7x10^{-30}$	0.98 (0.92-1.03)	0.4140	$3.0x10^{-30}$	-	-	-	-	-
GWAS	Psoriasis	rs1581803[c]	152592281	G/T	*LCE3B/LCE3D*	0.97 (0.90-1.04)	0.4396	1.22 (1.16-1.30)	$1.5x10^{-12}$	$1.6x10^{-12}$	-	-	-	-	-
GWAS	antagonistic	rs35722864	153040505	-/A	*SPRR* cluster	0.81 (0.75-0.88)	$1.0x10^{-7}$	1.13 (1.07-1.20)	$2.1x10^{-5}$	$4.8x10^{-13}$	0.85 (0.71-0.93)	0.0019	1.07 (1.01-1.14)	0.0211	$1.3x10^{-4}$

Chromsome 5q31.1

Data source	Effect	SNV	Position	Allele	Candidate Gene	Conditional Models[b]					Full Model				
						AE		Psoriasis		$P_{overall}$	AE		Psoriassis		$P_{overall}$
						OR (95% CI)	P	OR (95% CI)	P		OR (95% CI)	P	OR (95% CI)	P	
Ichip	antagonistic	rs6596086	131952222	C/T	*RAD50*	1.31 (1.22-1.41)	$1.7x10^{-13}$	0.86 (0.80-0.92)	$1.7x10^{-5}$	$5.7x10^{-21}$	1.17 (1.07-1.28)	$4.0x10^{-4}$	0.88 (0.81-0.96)	0.0023	$6.3x10^{-7}$
Ichip	AE	rs848	131996500	A/C	*IL13*	1.20 (1.10-1.23)	$5.6x10^{-5}$	0.96 (0.89-1.04)	0.3375	$4.1x10^{-5}$	1.12 (1.02-1.23)	0.0197	0.96 (0.88-1.05)	0.3515	0.0204
Ichip	AE	rs2299009	132042813	G/T	*IL4/KIF3A*	1.14 (1.06-1.23)	$7.9x10^{-4}$	0.99 (0.92-1.06)	0.7392	0.0018	1.16 (1.07-1.25)	$2.0x10^{-4}$	0.99 (0.92-1.06)	0.6657	$4.1x10^{-4}$
Ichip	AE	rs74458173	131621731	A/C	*SLC22A4*	1.57 (1.26-1.96)	$6.1x10^{-5}$	1.02 (0.80-1.30)	0.8590	$2.1x10^{-4}$	1.57 (1.26-1.96)	$5.7x10^{-5}$	1.02 (0.80-1.30)	0.8683	$2.0x10^{-4}$

Table 3.17.: Conditional analysis of the 1q21.3 and 5q31.1 regions showing disease-specific and opposing risk effects on AE and psoriasis.

[a] Conditional analysis of chr1q21.3 was conditioned on *FLG* for AE and *LCE3B/LCE3D* for psoriasis;

[b] stepwise conditional analysis at chr5q31.1 was carried out using multinomial regression models and resulted in three additional signals for AE; alternating colors (gray/white) indicate distinct LD blocks identified in stepwise analysis; this table shows only independent loci ($r^2 < 0.5$) and the SNV with the strongest association; the effect allele is underlined;

[c] rs1581803 tags the previously reported psoriasis SNV rs4112788 ($r^2 = 0.995$);

[d] $P_{overall}$ represents MNM P-value from the $2 - df$ linear hypothesis test for the overall SNV effect (see Table 2.3); Full model incorporates combined effects of independent SNVs identified by stepwise analyses.

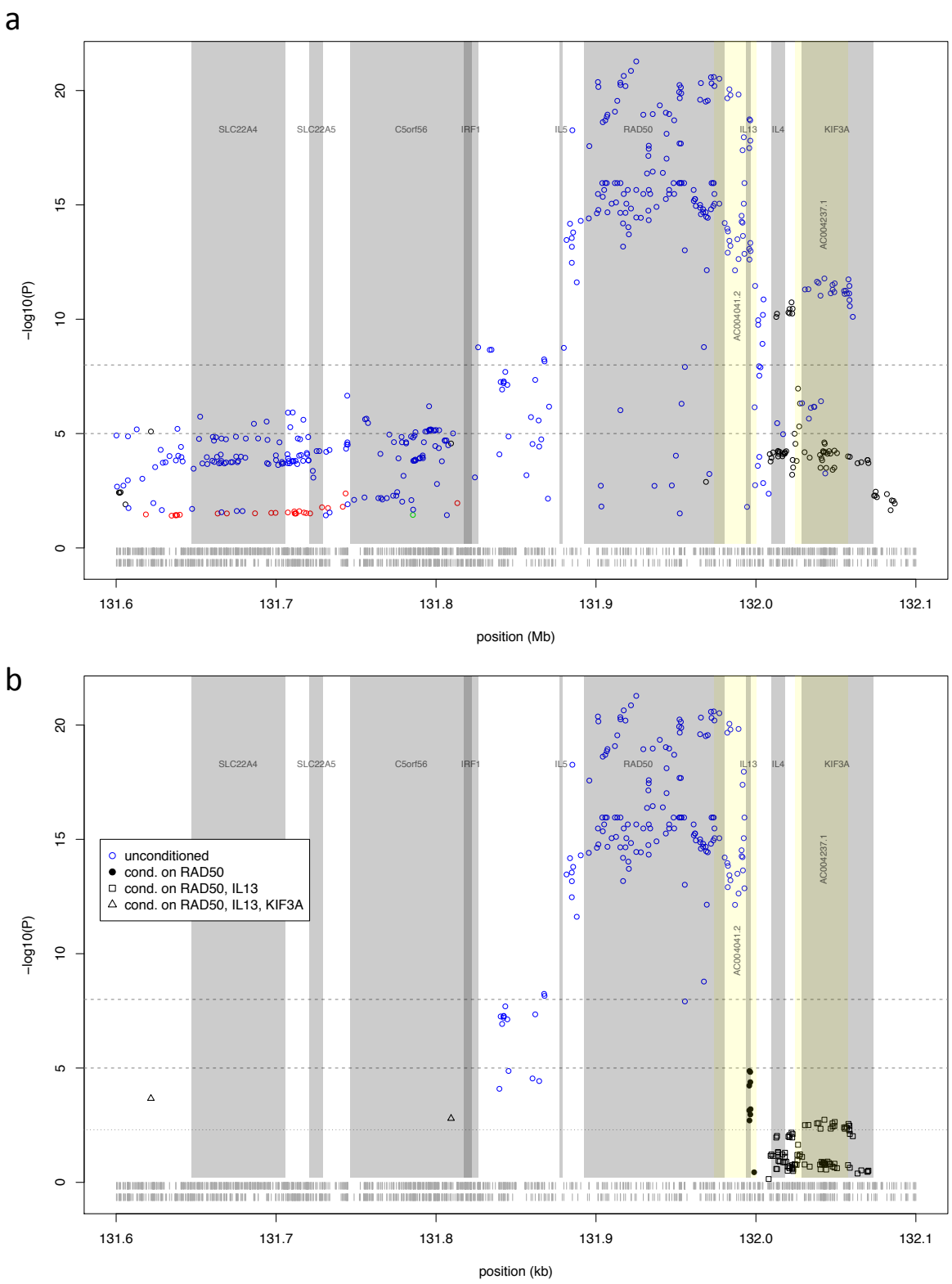


Figure 3.35.: Regional association within the cytokine cluster at 5q31.1.
(a) Multinomial regression model using GWAS and ImmunoChip data. **Black** circles indicate AE-specific, **red** psoriasis-specific, **blue** antagonistic and **green** agonistic SNV effects. (b) Conditional regional association plot of 5q31.1 by multinomial regression. Different symbols indicate association results after each step of analysis, as follows: Unconditioned results are shown by **blue** circles representing opposing effects in AE and psoriasis; **black** dots show AE-specific association results after conditioning on *RAD50*; **black** squares indicate AE-specific association after conditioning on *RAD50* and *IL13*; and **black** triangles indicate the AE-specific association after additionally conditioning on *KIF3A*. SNVs indicated by the same symbol are in LD. For both plots, vertical gray shading marks the positions of known genes (Genome Reference Consortium GRCh37), and horizontal dotted lines indicate significance thresholds of $P = 0.005$, 10^{-5} and 10^{-8}. The horizontal bands at the bottom indicate the coverage of the region by GWAS SNVs (upper row) and ImmunoChip SNVs (lower row).

3.6.2. Analysis of the MHC Confirms Multiple Psoriasis Loci and Identifies Opposing Loci for AE and Psoriasis

For an analysis of the extended HLA region, 21,913 SNVs selected by CCMA with $T_{max} > 4.68$ are taken forward to multinomial modeling. 18,114 SNVs are categorized as psoriasis-specific by both the CCMA and MNM. For conditional analyses the set of SNVs is reduced to those with consistent effect categorization by both methods and meeting the threshold of suggestive association in MNM ($P_{MNM} < 10^{-5}$). Then all SNVs in high LD ($r^2 > 0.8$ with the lead SNV) are discarded from the psoriasis-specific markers. Finally a set of 1,503 SNVs is obtained, which includes those SNVs previously reported for AE.[125]
Within the HLA region the strongest association with psoriasis is observed for the locus that tags the well-known psoriasis risk allele HLA-C*06:02.[106] The A allele of the lead SNV rs111576655 increases the risk for psoriasis by 3-fold ($OR_{PSO\ full} = 3.32$, $P_{PSO\ full} = 2.3 \times 10^{-69}$). Stepwise conditional analysis was carried out and after the final selection step a full model was calculated. The analysis reveals two additional independent loci contributing to the psoriasis risk at *MICA* (rs201374403, $OR_{PSO\ full} = 1.65$, $P_{PSO\ full} = 1.8 \times 10^{-25}$) and *HLA-A* (rs113573479, $OR_{PSO\ full} = 1.41$, $P_{PSO\ full} = 2.8 \times 10^{-17}$), as well as two loci with antagonistic effects at *HLA-DRB1* (rs28383201, $OR_{AE\ full} = 0.61$, $OR_{PSO\ full} = 1.18$, $P_{overall} = 6.5 \times 10^{-14}$) and *HLA-C* (rs1793889, $OR_{AE\ full} = 0.6$, $OR_{PSO\ full} = 1.18$, $P_{overall} = 1.1 \times 10^{-9}$) (Table 3.18, Figure 3.36).

Data source	Effect	SNV	Position	Allele	Candidate Gene	Conditional Models[a]					Full Model				
						AE		Psoriasis		$P_{overall}$	AE		Psoriassis		$P_{overall}$
						OR (95% CI)	P	OR (95% CI)	P		OR (95% CI)	P	OR (95% CI)	P	
Ichip	Psoriasis	rs111576655	31242731	A/T	**C*06:02**	0.84 (0.74-0.95)	0.0053	4.41 (4.10-4.74)	$3.1x10^{-376}$	$9.8x10^{-380}$	1.12 (0.81-1.54)	0.5071	3.32 (2.90-3.81)	$2.3x10^{-69}$	$3.2x10^{-65}$
Ichip	Psoriasis	rs201374403	31383754	T/TAG	*MICA*	0.78 (0.70-0.88)	$6.4x10^{-5}$	1.68 (1.56-1.80)	$7.9x10^{-48}$	$2.2x10^{-53}$	0.81 (0.67-0.96)	0.0174	1.65 (1.50-1.81)	$1.8x10^{-25}$	$1.0x10^{-26}$
Ichip	Psoriasis	rs113573479	29842444	A/G	*HLA-A*	0.89 (0.81-0.97)	0.0109	1.39 (1.30-1.49)	$6.6x10^{-25}$	$1.0x10^{-26}$	0.92 (0.81-1.04)	0.1948	1.41 (1.30-1.52)	$2.8x10^{-17}$	$2.7x10^{-17}$
Ichip	antagonistic	rs28383201	32574869	C/G	*HLA-DRB1*	0.59 (0.51-0.68)	$4.6x10^{-13}$	1.15 (1.06-1.24)	$4.5x10^{-4}$	$3.3x10^{-16}$	0.61 (0.52-0.71)	$3.4x10^{-10}$	1.18 (1.08-1.28)	$1.0x10^{-4}$	$6.5x10^{-14}$
Ichip	antagonistic	rs1793889	31222181	A/G	*HLA-C*	0.60 (0.50-0.73)	$2.5x10^{-7}$	1.18 (1.07-1.31)	0.0011	$1.1x10^{-9}$	0.60 (0.50-0.73)	$2.5x10^{-7}$	1.18 (1.07-1.31)	0.0011	$1.1x10^{-9}$

Table 3.18.: Conditional analysis of the MHC region on 6p21-22 showing psoriasis-specific and opposing risk effects in AD and psoriasis.

Effect allele is underlined; Table shows only independent loci ($r^2 < 0.5$) and the SNV with the strongest association.

[a] stepwise conditional analysis was carried out using multinomial regression models and resulted in three psoriasis-specific and two antagonistic signals.

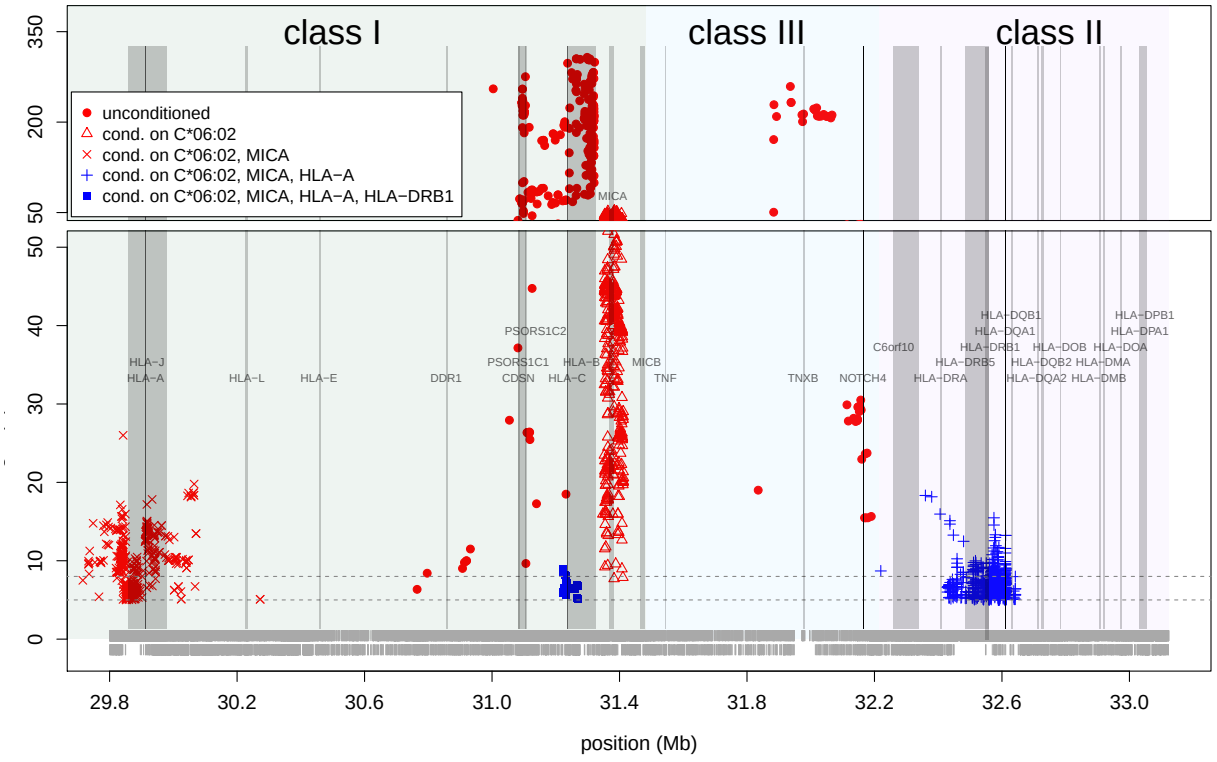


Figure 3.36.: Conditional regional association within the major histocompatibility complex at 6p21-22.

Symbols indicate association results after each step of analysis, as follows: Unconditioned psoriasis specific results are shown by **red** dots; **red** triangles show psoriasis specific association results after conditioning on **C*06:02** (known to be strongly associated with psoriasis); **red** 'X's indicate psoriasis specific association results after conditioning on **C*06:02** and *MICA*; **blue** '+'s indicate the association results after conditioning on **C*06:02**, *MICA* and *HLA-A* with antagonistic effects on AE and psoriasis; **blue** squares indicate the residual association after conditioning on **C*06:02**, *MICA*, *HLA-A* and *HLA-DRB1* with antagonistic effects on AE and psoriasis. SNVs indicated by the same symbol are in LD. Vertical shading marks the positions of known genes (identified from the UCSC Genome Browser GRCh37/hg19 accessed Feb. 2009) and HLA classes; horizontal dotted lines indicate significance thresholds of $P = 10^{-5}$ and 10^{-8}. The horizontal bands at the bottom indicate the coverage of the region by GWAS SNVs (upper row) and ImmunoChip SNVs (lower row).

4. Discussion

The field of genetic epidemiology has changed rapidly over the last decade, mainly driven by improvement of genotyping technology. The GWAS era since the year 2005 has been very successful and contributed strongly to the understanding of the genetic architecture underlying complex diseases.[40] Moreover, GWAS have identified a remarkable overlap of genetic factors among complex diseases[1,4] and revealed many shared loci across immune-mediated disorders.[126] Recently, the focus in genetic epidemiology has been drawn to identify pleiotropic loci, and several methods have been proposed for this purpose.[6,7]
Meta-analysis is a statistical tool to combine different studies while increasing power. Here, a meta-analysis based method, the **C**ombined & **C**ontrast **M**eta-**A**nalysis (CCMA), is proposed, which we have recently shown to be a useful approach for identifying pleiotropic loci for atopic eczema and psoriasis.[106] In this work the null distribution of the CCMA test statistic has been determined, which corresponds to an exponential distribution. It has been shown that CCMA has comparable power for detecting disease specific, agonistic and antagonistic effects to the frequently used Subset-Based Meta-Analysis.[7] The CCMA method is easy to implement and allows a fast pre-filtering of SNVs by quantifying agonistic and antagonistic effects. Modifications of the method turned out to increase the power for finding antagonistic effects. It is advisable to choose the appropriate version of the test statistic prior to the analysis.
The CCMA method can also be applied to z-scores from two single studies, one for each trait of interest, and it does not rely on the availability of results from meta-analyses. Moreover, in case of using results from meta-analyses it is possible to choose random effects models for estimating T_i by applying the DL-method or any proposed refinement of the variance estimate (e.g. $\mathrm{Var}(\hat{\theta}_{HS})$).[58,59]
A confidence interval for the CCMA test statistic can be derived which corresponds well with the significance level of the hypothesis test. By inserting the four test statistics and solving for the numerator a confidence interval for the effect of interest can be obtained. However, apart from the disease specific effect the interpretation of the confidence interval for agonistic and antagonistic effects is not straightforward and only approximate confidence intervals for the sum or the difference can be deduced. In general, an appropriate hypothesis test can be specified for each confidence interval, while the reverse is not always possible.[108] Further, it remains to be shown whether the CCMA method can be extended for comparing and contrasting more than two traits, as is the case for the subset-based meta-analysis.[7]
The variants obtained by CCMA can be further investigated using e.g. multinomial regression models, replication studies or functional analysis. Of note, the CCMA method allows researchers to infer the mode of pleiotropy directly by looking at which of the four constituent statistics T_1, T_2, $T_{12\mathrm{agonistic}}$ and $T_{12\mathrm{antagonistic}}$ is maximal. The method

shows good agreement with results from the multinomial regression model, which uses the complete information of individual genotype data. Moreover, CCMA outperforms multinomial regression models in terms of computing time by far and it has the useful property of using only summary statistics from different studies. Therefore, individual genotype data do not need to be transferred to a central site of meta-analysis.
In comparison to the previously proposed COMBINED & OVERLAP method,[6] CCMA allows researchers to identify disease specific effects, it does not rely on elaborate selection criteria, and commonly used thresholds of *suggestive* and *genomewide* significant association can be directly applied. Furthermore, the analysis is less complicated as no allele-recoding has to be done for identifying opposing effects on both diseases.
Finally, the CCMA method, which is calculated in a straightforward way, can also be applied to other genome-wide molecular data (e.g. expression, epigenomics, metabolomics) as well as to other research questions such as those encountered in environmental epidemiology.
From a medical point of view, contrasting atopic eczema and psoriasis might not be the most interesting scientific topic at first sight, since both diseases rarely co-occur within the same patient and thus not much genetic overlap is expected. However, both are common chronic inflammatory skin diseases with epidermal and immunological abnormalities. Interestingly, the genome-wide comparative analysis performed here confirms high levels of genomic co-incidence between AE and psoriasis, indicating that common molecular mechanisms are involved. This is in concordance with the pivotal role of epidermal barrier defects and T-cell dominated inflammation in both diseases.[127] The majority of genetic regions identified for AE are localized in close proximity to peaks that have been mapped for psoriasis;[128] for example, an atopic eczema locus on chromosome 1q21.3 within the EDC lies within 30 cM of a genome-wide significant psoriasis locus.[129]
We could identify six antagonistic loci, i.e. exerting opposing effects on AE and psoriasis, that meet the genome-wide significance threshold within the EDC (1q21.3), the cytokine cluster (5q31.1), the HLA-region (6p21-22), *PRKRA* (2q31.2) and *ANXA6/TNIP1* (5q33.1). These loci include genes predominantly contributing to the development and maintenance of the epidermal barrier and immunological response. Of note, all identified loci display antagonistic effects, which agrees with the epidemiological observation of lower-than-expected coincidence between both diseases,[130] and two of them (*ANXA6/TNIP1*, *PRKRA*) have not previously been reported to be associated with psoriasis and/or AE.
The antagonistic effect of variation in *PRKRA* may arise from its involvement in cellular response to environmental stress and thus may reflect the striking differential susceptibility to viral and bacterial skin infections observed in AE and psoriasis. The opposing effect of variation in *ANXA6* indicates a role for calcium-dependent effects in defining patterns of skin inflammation.[106]
On chromosome 1q21.3, FLG-AS1 is a plausible candidate to induce antagonistic effects on AE and psoriasis via transcriptional and post-transcriptional regulation of gene expression, while we hypothesize that *RAD50* on chromosome 5q31.1 may mediate the differential risk for AE and psoriasis by variation in DNA repair leading to a different Th2 response.[106]
The MHC analysis confirms several previously reported loci within HLA class I with independent effects on psoriasis, of which the strongest is represented by SNVs tagging

the classical HLA allele Cw*06:02.[131] The susceptibility of primarily HLA class I loci to psoriasis may mirror the psoriasis-associated (auto-)antigen presentation to pathogenic CD8+ T cells,[132] which show increased levels in lesional psoriatic skin. Increased levels of CD8+ T cells are also found in lesional AE skin with however strikingly different cytokine profiles compared to psoriasis,[133] which may be corroborated by the identified opposing locus at *HLA-C*. The antagonistic locus at HLA class II for AE and psoriasis may reflect the differential responses to pathogenic and allergenic peptides presented to CD4+ T cells.[134]

Generally, the comparative analysis of AE and psoriasis demonstrates that the primary genetic factors underlying AE influence the skin barrier function, whilst the factors responsible for psoriasis are involved in (auto-)antigen recognition. Several pleiotropic loci with antagonistic effects contribute to opposing mechanisms of adaptive immunity in both AE and psoriasis.[106]

In summary, our method is an appealing approach to screen for shared and disease specific genetic loci and to dissect cross-phenotype association using available GWAS data. It can also serve as a good starting point to investigate closely related disorders, such as atopic eczema and asthma or psoriasis vulgaris and psoriatric arthritis, to identify shared and unique genetic risk loci. These loci may represent overlapping pathophysiological mechanisms and important switch points in pathways determining susceptibility to one or both diseases. Characterization of agonistic and antagonistic molecular mechanisms across complex phenotypes will enrich our knowledge of biology and disease, which provides a basis for improved or newly developed therapies.

A. Supplement

A.1. Calculation of the Subset-Based Meta-Analysis P-value in the case of two traits

In the case of two traits formula (2.6) can be written as

$$\begin{aligned}\tilde{P}_{\mathrm{DLM}} = &\int_T^\infty 2 \cdot Pr\left(|Z(12)| < z \,|Z(1) = z\right) \phi(z)\, dz \\ &+ \int_T^\infty 2 \cdot Pr\left(|Z(12)| < z \,|Z(2) = z\right) \phi(z)\, dz \\ &+ \int_T^\infty 2 \cdot Pr\left(|Z(1)| < z \,|Z(12) = z\right) \cdot Pr\left(|Z(2)| < z \,|Z(12) = z\right) \phi(z)\, dz \end{aligned} \tag{A.1}$$

According to Bhattacharjee and colleagues[7] $Z(S+k) < z$ can be substituted by $Z(k) < u_k(z)$ for the upper bound and $Z(S-k) < z$ by $Z(k) > l_k(z)$ for the lower bound. In order to derive upper and lower bound, they defined $Z(S \pm k) = w_S Z(S) \pm w_k Z(k)$. Then conditioned on $Z(S) = z$, $Z(S \pm k) < z$ can be rewritten by

$$\begin{aligned} Z(S+k) &< z \\ w_S\, z + w_k\, Z(k) &< z \\ Z(k) &< \frac{(1-w_S)}{w_k} z \equiv u_k(z) \end{aligned}$$

$$\begin{aligned} Z(S-k) &< z \\ w_S\, z - w_k\, Z(k) &< z \\ Z(k) &> \frac{(w_S-1)}{w_k} z \equiv l_k(z) \end{aligned} \tag{A.2}$$

where w_S and w_k are defined as

$$w_k = \sqrt{\pi_k(S+k)} = \sqrt{\frac{N_k}{\sum_{s\in(S+k)} N_s}}$$

$$w_S = \sqrt{\pi_S(S+k)} = \sqrt{\frac{\sum_{s\in S} N_s}{\sum_{s\in(S+k)} N_s}}$$

The similarity of $\pi_k(S+k)$ and $\pi_S(S+k)$ to $\pi_k(S)$ in eq. (2.4) is obvious. With eq. (A.2) formula (A.1) can be written in an analogous way to eq. (2.7) as

$$\begin{aligned}
\tilde{P}_{\mathrm{DLM}} &= \int_T^\infty 2 \cdot Pr\left(l_2(z) < Z(2) < u_2(z) | Z(1) = z\right) \phi(z)\, dz \\
&+ \int_T^\infty 2 \cdot Pr\left(l_1(z) < Z(1) < u_1(z) | Z(2) = z\right) \phi(z)\, dz \\
&+ \int_T^\infty 2 \cdot Pr\left(l_2(z) < Z(2) < u_2(z) | Z(12) = z\right) \cdot Pr\left(l_1(z) < Z(1) < u_1(z) | Z(12) = z\right) \phi(z)\, dz \\
&= \int_T^\infty 2 \cdot \Omega_2\, \phi(z)\, dz \\
&+ \int_T^\infty 2 \cdot \Omega_1\, \phi(z)\, dz \\
&+ \int_T^\infty 2 \cdot \Omega_2 \cdot \Omega_1\, \phi(z)\, dz
\end{aligned} \tag{A.3}$$

with Ω_k denoting the weights determined by each neighboring subset.
Bhattacharjee and colleagues[7] argue that under the null hypothesis H_0 "the vector of test statistics $Z(S)$ for different values of S should follow a multivariate normal distribution with a mean of zero and unit variance for each component".[7] This allows the evaluation of each weight Ω_k in eq. (A.3) by a conditional univariate normal distribution, which can be derived by using the formula given in Fahrmeir and Hamerle[107]

$$Z_A | Z_B = z_B \sim N(\mu_A + \frac{\sigma_A}{\sigma_B} \rho\, (z_B - \mu_B),\ (1 - \rho^2)\, \sigma_A^2) \tag{A.4}$$

With the assumption of a multivariate standard normal distribution under H_0, the conditional univariate normal distribution in eq. (A.4) can be reduced to

$$Z_A | Z_B = z_B \sim N(\rho\, z_B,\ 1 - \rho^2) \tag{A.5}$$

In order to calculate the weights Ω_k by conditional univariate normal distributions the correlation ρ between a pair of subsets has to be determined. In the case of independent studies and assuming that the statistics $Z(\cdot)$ are asymptotically standard normal distributed, the authors defined the correlation ρ as[7]

$$\begin{aligned}
\rho = \sigma_{A,B} &= \mathrm{Cov}\{Z(A), Z(B)\} \\
&= \sum_{k \in A,B} \sqrt{\pi_k(A)} \sqrt{\pi_k(B)} \\
&= \sum_{k \in A,B} \sqrt{\frac{N_k}{\sum_{k \in A} N_k}} \sqrt{\frac{N_k}{\sum_{k \in B} N_k}}
\end{aligned} \tag{A.6}$$

For calculating the weights Ω_k in the case of two traits, Z_A and Z_B are replaced by $Z(1)$, $Z(2)$ and $Z(12)$, and $\sigma_{A,B}$ by $\sigma_{1,12}$, $\sigma_{2,12}$, $\sigma_{12,1}$ or $\sigma_{12,2}$. Applying eq. (A.6) it turns out that $\sigma_{1,12} = \sigma_{12,1} = w_1$ and $\sigma_{2,12} = \sigma_{12,2} = w_2$:

$$\begin{aligned}\sigma_{1,12} &= \sqrt{\frac{N_1}{N_1}}\sqrt{\frac{N_1}{N_1+N_2}} = \sqrt{\frac{N_1}{N}} = w_1 \\ \sigma_{2,12} &= \sqrt{\frac{N_2}{N_2}}\sqrt{\frac{N_2}{N_1+N_2}} = \sqrt{\frac{N_2}{N}} = w_2\end{aligned} \tag{A.7}$$

Replacing the weights Ω_k by the difference of conditional normal probabilities between the upper and the lower bounds in eq. (A.3) yields

$$\begin{aligned}\tilde{P}_{\mathrm{DLM}} = &\int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{u_2(z) - w_1\, z}{\sqrt{1-w_1^2}}\right) - \Phi\left(\frac{l_2(z) - w_1\, z}{\sqrt{1-w_1^2}}\right)\right\} \phi(z)dz \\ &+ \int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{u_1(z) - w_2\, z}{\sqrt{1-w_2^2}}\right) - \Phi\left(\frac{l_1(z) - w_2\, z}{\sqrt{1-w_2^2}}\right)\right\} \phi(z)dz \\ &+ \int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{u_2(z) - w_1\, z}{\sqrt{1-w_1^2}}\right) - \Phi\left(\frac{l_2(z) - w_1\, z}{\sqrt{1-w_1^2}}\right)\right\} \cdot \\ &\qquad \left\{ \Phi\left(\frac{u_1(z) - w_2\, z}{\sqrt{1-w_2^2}}\right) - \Phi\left(\frac{l_1(z) - w_2\, z}{\sqrt{1-w_2^2}}\right)\right\} \phi(z)dz \\ \\ = &\int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{(1-w_1)\, z - w_1 w_2 z}{w_2\sqrt{1-w_1^2}}\right) - \Phi\left(\frac{(w_1-1)\, z - w_1 w_2 z}{w_2\sqrt{1-w_1^2}}\right)\right\} \phi(z)dz \\ &+ \int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{(1-w_2)\, z - w_1 w_2 z}{w_1\sqrt{1-w_2^2}}\right) - \Phi\left(\frac{(w_2-1)\, z - w_1 w_2 z}{w_1\sqrt{1-w_2^2}}\right)\right\} \phi(z)dz \\ &+ \int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{(1-w_1) z - w_1 w_2 z}{w_2\sqrt{1-w_1^2}}\right) - \Phi\left(\frac{(w_1-1) z - w_1 w_2 z}{w_2\sqrt{1-w_1^2}}\right)\right\} \cdot \\ &\qquad \left\{ \Phi\left(\frac{(1-w_2) z - w_1 w_2 z}{w_1\sqrt{1-w_2^2}}\right) - \Phi\left(\frac{(w_2-1) z - w_1 w_2 z}{w_1\sqrt{1-w_2^2}}\right)\right\} \phi(z)dz\end{aligned}$$

with $\Phi(\cdot)$ denoting the cumulative standard normal distribution function.

From (A.7) it can be seen that $w_1^2 = 1 - w_2^2$ and $w_2^2 = 1 - w_1^2$, above formula can be simplified to

$$\begin{aligned}
\tilde{P}_{\text{DLM}} = &\int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{(1-w_1)\,z - w_1 w_2 z}{w_2^2} \right) - \Phi\left(\frac{(w_1-1)\,z - w_1 w_2 z}{w_2^2} \right) \right\} \phi(z) dz \\
&+ \int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{(1-w_2)\,z - w_1 w_2 z}{w_1^2} \right) - \Phi\left(\frac{(w_2-1)\,z - w_1 w_2 z}{w_1^2} \right) \right\} \phi(z) dz \\
&+ \int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{(1-w_1)z - w_1 w_2 z}{w_2^2} \right) - \Phi\left(\frac{(w_1-1)z - w_1 w_2 z}{w_2^2} \right) \right\} \cdot \\
&\qquad \left\{ \Phi\left(\frac{(1-w_2)z - w_1 w_2 z}{w_1^2} \right) - \Phi\left(\frac{(w_2-1)z - w_1 w_2 z}{w_1^2} \right) \right\} \phi(z) dz \\
\\
= &\int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{(1-(1+w_2)w_1)z}{w_2^2} \right) - \Phi\left(\frac{(w_1(1-w_2)-1)z}{w_2^2} \right) \right\} \phi(z) dz \\
&+ \int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{(1-(1+w_1)w_2)z}{w_1^2} \right) - \Phi\left(\frac{(w_2(1-w_1)-1)z}{w_1^2} \right) \right\} \phi(z) dz \\
&+ \int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{(1-(1+w_2)w_1)z}{w_2^2} \right) - \Phi\left(\frac{(w_1(1-w_2)-1)z}{w_2^2} \right) \right\} \cdot \\
&\qquad \left\{ \Phi\left(\frac{(1-(1+w_1)w_2)z}{w_1^2} \right) - \Phi\left(\frac{(w_2(1-w_1)-1)z}{w_1^2} \right) \right\} \phi(z) dz
\end{aligned}$$

Applying some basic transformations and replacing w_1 and w_2 by $\sqrt{N_1/N}$ and $\sqrt{N_2/N}$ yields

$$\begin{aligned}
\tilde{P}_{\text{DLM}} = &\int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{\left(1 - \left(1 + \sqrt{\frac{N_2}{N}}\right)\sqrt{\frac{N_1}{N}}\right) z}{\frac{N_2}{N}} \right) - \Phi\left(\frac{\left(\sqrt{\frac{N_1}{N}}\left(1 - \sqrt{\frac{N_2}{N}}\right) - 1\right) z}{\frac{N_2}{N}} \right) \right\} \phi(z) dz \\
&+ \int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{\left(1 - \left(1 + \sqrt{\frac{N_1}{N}}\right)\sqrt{\frac{N_2}{N}}\right) z}{\frac{N_1}{N}} \right) - \Phi\left(\frac{\left(\sqrt{\frac{N_2}{N}}\left(1 - \sqrt{\frac{N_1}{N}}\right) - 1\right) z}{\frac{N_1}{N}} \right) \right\} \phi(z) dz \\
&+ \int_T^\infty 2 \cdot \left\{ \Phi\left(\frac{\left(1 - \left(1 + \sqrt{\frac{N_2}{N}}\right)\sqrt{\frac{N_1}{N}}\right) z}{\frac{N_2}{N}} \right) - \Phi\left(\frac{\left(\sqrt{\frac{N_1}{N}}\left(1 - \sqrt{\frac{N_2}{N}}\right) - 1\right) z}{\frac{N_2}{N}} \right) \right\} \cdot \\
&\qquad \left\{ \Phi\left(\frac{\left(1 - \left(1 + \sqrt{\frac{N_1}{N}}\right)\sqrt{\frac{N_2}{N}}\right) z}{\frac{N_1}{N}} \right) - \Phi\left(\frac{\left(\sqrt{\frac{N_2}{N}}\left(1 - \sqrt{\frac{N_1}{N}}\right) - 1\right) z}{\frac{N_1}{N}} \right) \right\} \phi(z) dz
\end{aligned}$$

$$
\begin{aligned}
\tilde{P}_{\mathrm{DLM}} = \int_T^\infty 2 \cdot \Bigg\{ & \Phi\left(\frac{\left(N - \sqrt{NN_1}\left(\sqrt{\frac{N_2}{N}} + 1\right)\right) z}{N_2} \right) - \\
& \Phi\left(\frac{-\left(\sqrt{NN_1}\left(\sqrt{\frac{N_2}{N}} - 1\right) + N\right) z}{N_2} \right) \Bigg\} \phi(z) dz \\
+ \int_T^\infty 2 \cdot \Bigg\{ & \Phi\left(\frac{\left(N - \sqrt{NN_2}\left(\sqrt{\frac{N_1}{N}} + 1\right)\right) z}{N_1} \right) - \\
& \Phi\left(\frac{-\left(\sqrt{NN_2}\left(\sqrt{\frac{N_1}{N}} - 1\right) + N\right) z}{N_1} \right) \Bigg\} \phi(z) dz \\
+ \int_T^\infty 2 \cdot \Bigg\{ & \Phi\left(\frac{\left(N - \sqrt{NN_1}\left(\sqrt{\frac{N_2}{N}} + 1\right)\right) z}{N_2} \right) - \\
& \Phi\left(\frac{-\left(\sqrt{NN_1}\left(\sqrt{\frac{N_2}{N}} - 1\right) + N\right) z}{N_2} \right) \Bigg\} \cdot \\
& \Bigg\{ \Phi\left(\frac{\left(N - \sqrt{NN_2}\left(\sqrt{\frac{N_1}{N}} + 1\right)\right) z}{N_1} \right) - \\
& \Phi\left(\frac{-\left(\sqrt{NN_2}\left(\sqrt{\frac{N_1}{N}} - 1\right) + N\right) z}{N_1} \right) \Bigg\} \phi(z) dz
\end{aligned}
$$

A.2. Study Subjects

Genome-wide genotype data were obtained for samples from three case-control cohorts for AE and psoriasis, respectively, totalling 2,262 AE and 4,489 psoriasis cases and 12,333 controls. The German AE patients (panels A and B in Table A.1) were recruited from tertiary dermatology clinics at Munich, as part of the GENEVA study, as well as from the University of Kiel, the University of Bonn and the Charité Universitätsmedizin in Berlin. AE was diagnosed by experienced dermatologists according to the UK Diagnostic Criteria.[135] German controls were obtained from the PopGen biorepository,[136] the population-based KORA study in southern Germany[137] and the German part of ISAAC II.[138] The Irish AE case collection was recruited from the secondary and tertiary pediatric dermatology clinic at Our Lady's Children's Hospital, Crumlin, Dublin (panel C).

Irish population control samples were obtained from healthy adult blood donors as part of the Trinity Biobank.[139] The German psoriasis patients were recruited from the tertiary dermatology clinic at the University of Kiel, and German controls were again obtained from the PopGen biorepository as well as the KORA study (panel D) with the subset of KORA participants being independent from the ones used as controls for AE. The British psoriasis case-control study is part of the Wellcome Trust Case Control Consortium[140] (panel E). Cases and controls of the US psoriasis study (CASP) were recruited at dermatology clinics at the Universities of Utah, Michigan, Texas and California, as described elsewhere[131] (panel F).
Genotype data from the customized Immunochip array were obtained for samples from previous studies[2,3] (see Table A.1).

A.3. Quality Control & Genomewide Genotype Imputation

Prior to imputation, quality control (QC) and standard GWAS analysis of genotyped SNPs was carried out using PLINK[115] and R.[111] In brief, we excluded samples with extensive missing data rate (>5%), excess of heterozygosity or homozygosity, and discrepant gender determined on the basis of average X-chromosomal heterozygosity compared to the gender recorded in the database. We then examined identity-by-state (IBS) sharing and estimated identity-by-descent (IBD) on a pruned SNP set (i.e., with SNVs being independent in terms of LD; $r^2 < 0.5$) between all pairs of individuals and deleted resulting duplicates or closely related samples with estimated IBD>0.1875 (i.e., halfway between expected IBD for second- and third degree relatives). Multidimensional scaling (MDS) of the pairwise IBS matrix was carried out to identify and delete outliers of unusual ancestry and to calculate genome-wide principal component scores for each individual (Figure A.1). SNVs showing a missing rate of >5%, deviation of Hardy-Weinberg equilibrium $P_{HWE} < 10^{-8}$ or minor allele frequency MAF<5% were discarded.
After QC, the resulting SNVs and samples were analyzed for association with the phenotype applying logistic regression using age, sex and the first four principal components as covariates. Results from each study panel were investigated to determine whether established GWAS loci were identified for the respective trait of interest, and genomic control (GC)[141] inflation factors were calculated. The GC method is commonly used in GWAS and calculates an inflation factor

$$\lambda = \frac{median(X^2)}{median(\chi^2(1))} \tag{A.8}$$

where X^2 are the observed test statistics for each SNP, which asymptotically follow a $\chi^2(1)$-distribution if no population stratification is present. Then the observed test statistics are divided by λ in order to correct for population stratification

$$X^2_\lambda = \frac{X^2}{\lambda}$$

Panel	Numbers		Collection	Platform
GWAS data				
A	Atopic eczema	663	Munich / Bonn	Illumina 300k
Germany	Controls	786	PopGen / ISAAC	
B	Atopic eczema	993	Munich / Kiel / Berlin	Affymetrix 6.0
Germany	Controls	1513	KORA	
C	Atopic eczema	606	Dublin / Dundee	Illumina 610k
Ireland	Controls	1794	TRINITY (Dublin)	Affymetrix 6.0
D	Psoriasis	492	Kiel	Illumina 550k
Germany	Controls	1161	PopGen / KORA	
E	Psoriasis	2622	WTCCC2	Illumina 1M
United Kingdom	Controls	5667		
F	Psoriasis	1375	CASP	Perlegen
United States	Controls	1412		
Immunochip data				
Germany	Atopic eczema	2425	Munich / Bonn / Berlin	Immunochip
	Psoriasis	572	Kiel	
	Controls	5449	PopGen / KORA / HNR	
			Munich / Berlin / Bonn	
United States	Psoriasis	1351	UMich / NPH / HFH	
	Controls	2694	UMich / FIMR / NPH	
Canada	Psoriasis	362	UMich / NPH / HFH	
	Controls	20	UToronto	
Estonia	Psoriasis	1295	UTartu / EGCUT	
	Controls	898	EGCUT	

Table A.1.: Case control collection for atopic eczema and psoriasis

PopGen, PopGen biorepository; ISAAC, International Study of Asthma and Allergies in Childhood; TRINITY, Trinity Biobank Controls, Dublin, Ireland; WTCCC2, Wellcome Trust Case Control Consortium 2; HNR, Heinz Nixdorf Recall study; CASP, Collaborative Association Study of Psoriasis; UMich, University of Michigan; HFH, Henry Ford Hospital, NPF, National Psoriasis Foundation Victor Henschel BioBank; FIMR, The Feinstein Institute for Medical Research; UToronto, University of Toronto; MU, Memorial Hospital Newfoundland; UTartu, University of Tartu; EGCUT, Estonian Genome Center University of Tartu.

and the GC-corrected P-value is calculated by

$$P = 1 - F_1(X_\lambda^2).$$

where F_1 is the distribution function of the $\chi^2(1)$ distribution.
Any SNVs showing significant association were checked (e.g. by visual inspection of the intensity cluster plots and investigation of consistency of LD with surrounding markers) and those SNVs deemed unreliable were removed. The final data sets of high quality SNVs were pre-phased using SHAPEIT[36] and subsequently imputed using IMPUTE2[35] using

the 1000 genomes reference panel (integrated variant set, release March 2012).[20] Post imputation SNVs with low imputation quality (info score<0.4), call rate<95%, deviation from $P_{HWE} < 10^{-8}$ or MAF<5% were excluded. A final data set of approximately 5.2 million SNVs in 2,079 AE cases and 3,867 controls as well as 4,212 psoriasis cases and 8,032 controls were eligible for subsequent analysis.
Association testing for each study was carried out with SNPTEST[35] using a frequentist approach (i.e., drawing conclusions only from sample data in contrast to Bayesian statistics, which also uses prior information) with allele dosages (option `-method expected`). Meta-analysis for each disease was performed using METAL[100] based on the fixed effects model as described in Section 2.1 using GC (A.8) to adjust for population stratification on the meta-analysis level.

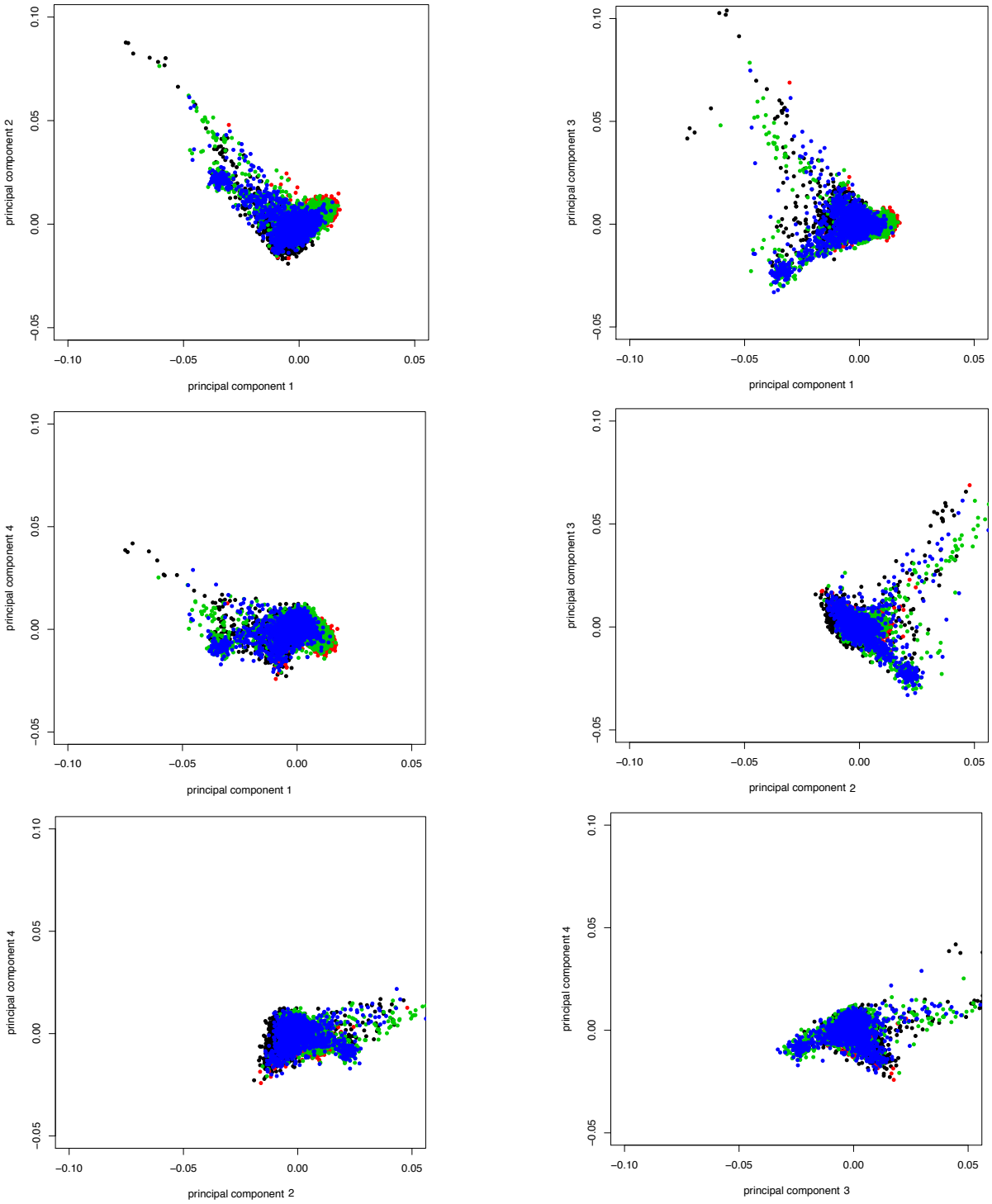


Figure A.1.: Multidimensional scaling of the IBS matrix based on 15,000 independent SNVs reveals the first 4 principal component scores, which are plotted against each other, describing the genetic distance between individuals. Color code indicate different ethnicities: **black=German**, **red=Irish**, **green=UK** and **blue=US**.

Lead SNV	Chr	Position	T_{max}	# clumped SNVs	Effect	Gene
chr1:8367013:D	1	8367013	4.700	0	AGONIST	-
rs2103876	1	12053100	5.107	24	AGONIST	MFN2
rs2076346	1	24083649	5.030	0	PSO	-
rs7534674	1	67646463	4.729	0	PSO	C1orf141
chr1:67679939:I	1	67679939	4.978	2	PSO	IL23R
rs11209026	1	67705958	5.544	2	PSO	IL23R
chr1:67708996:I	1	67708996	4.823	1	PSO	-
rs11209033	1	67744500	5.733	39	PSO	-
rs6683107	1	152117706	4.754	2	AE	-
rs10788823	1	152161648	5.876	1	AE	FLG-AS1
rs35058813	1	152168717	5.709	11	ANTAGONIST	FLG-AS1
rs4845425	1	152195034	6.078	12	ANTAGONIST	FLG-AS1
chr1:152209467:D	1	152209467	5.823	0	AE	FLG-AS1
rs12046937	1	152229109	5.402	2	AE	FLG-AS1
rs2146118	1	152412196	7.052	20	AE	-
chr1:152440012:I	1	152440012	7.775	41	AE	-
rs12081541	1	152441366	6.499	3	AE	-
chr1:152456562:D	1	152456562	7.547	0	AE	-
rs11205012	1	152478813	5.451	0	AE	LCE5A
rs12116609	1	152497685	8.579	163	AE	-
rs138038303	1	152521123	5.113	1	AE	-
rs17659389	1	152539393	4.925	0	ANTAGONIST	LCE3E
rs138035792	1	152588078	7.239	33	PSO	LCE3B
rs11587218	1	152603091	5.769	183	PSO	-
chr1:152706315:I	1	152706315	5.126	4	PSO	-
chr1:152757094:D	1	152757094	7.210	26	AE	LCE1E
rs2065206	1	152771293	5.599	6	ANTAGONIST	LCE1D
rs11586313	1	152890470	6.091	32	ANTAGONIST	-
rs11586024	1	152941933	4.732	0	ANTAGONIST	SPRR4
rs10788854	1	153002320	6.132	224	ANTAGONIST	SPRR1B
rs3120745	1	153019258	5.648	0	ANTAGONIST	SPRR2D
rs184299	1	153080515	5.725	7	ANTAGONIST	SPRR2E
rs10802516	1	247653412	4.760	0	ANTAGONIST	OR2W5
chr2:60979663:I	2	60979663	4.952	0	PSO	PAPOLG
rs1177213	2	61079090	6.657	29	PSO	-
rs10208309	2	61178509	5.180	4	AGONIST	PUS10
rs2111485	2	163110536	5.199	1	PSO	-
rs17716942	2	163260691	5.917	0	PSO	KCNH7
chr2:179300971	2	179300971	8.781	2	PSO	PRKRA
rs9406386	2	179315757	5.145	0	AGONIST	PRKRA
rs896056	3	13583528	5.365	0	AE	FBLN2
rs2455826	3	15663060	5.663	0	PSO	BTD
rs854216	3	45379777	4.777	0	PSO	-
rs7687602	4	107849340	4.824	0	AE	DKK2
rs11727622	4	131723791	4.867	6	AGONIST	-
chr5:96101066:I	5	96101066	5.271	12	PSO	CAST
rs39841	5	96120170	5.758	28	PSO	ERAP1
rs17166050	5	131915213	5.990	97	ANTAGONIST	RAD50
rs847	5	131996669	7.282	3	ANTAGONIST	IL13
rs1295683	5	131998876	4.780	0	ANTAGONIST	IL13
rs2243208	5	132001151	5.108	1	ANTAGONIST	IL13
chr5:132007974:D	5	132007974	4.968	0	ANTAGONIST	IL4
rs4705962	5	132028858	4.918	7	AE	KIF3A
rs115682037	5	150464017	6.499	1	PSO	TNIP1

Lead SNV	Chr	Position	T_{max}	# clumped SNVs	Effect	Gene
rs1024995	5	150476004	5.428	0	PSO	TNIP1
rs1421898	5	158111073	5.064	0	PSO	-
rs4444955	5	158564055	5.006	0	PSO	-
rs10077858	5	158614087	6.106	16	PSO	RNF145
rs57720492	5	158628703	6.255	5	PSO	RNF145
rs181071938	5	158636512	5.297	0	PSO	RNF145
rs254847	5	158670763	5.750	13	PSO	-
chr5:158685904:I	5	158685904	5.247	33	PSO	UBLCP1
rs7730390	5	158730792	9.156	31	PSO	-
rs2853694	5	158749088	9.512	4	PSO	IL12B
rs11135059	5	158771337	10.012	27	PSO	-
rs12153168	5	158785885	4.838	1	PSO	-
rs12657636	5	158810793	6.118	1	PSO	-
rs10046001	5	158814577	5.622	12	PSO	-
chr5:158826872:I	5	158826872	9.648	1	PSO	-
rs11743870	5	158849200	6.567	37	PSO	-
rs10515802	5	158860636	9.872	20	PSO	-
rs2161417	5	158862238	8.449	0	PSO	-
rs12188351	5	168386089	5.388	0	PSO	SLIT3
rs12202284	6	471136	4.896	0	PSO	-
rs2476847	6	549390	4.953	1	PSO	EXOC2
rs9504361	6	577820	5.122	0	PSO	EXOC2
chr6:111460283:I	6	111460283	5.385	1	PSO	SLC16A10
rs3912092	6	111572812	6.337	13	PSO	-
rs240976	6	111607992	5.961	84	PSO	-
rs465969	6	111655530	6.840	0	PSO	REV3L
rs240991	6	111668174	7.677	6	PSO	REV3L
rs463242	6	111668239	5.784	93	PSO	REV3L
chr6:111789955:D	6	111789955	5.160	0	PSO	REV3L
rs9320367	6	111897852	5.857	11	PSO	TRAF3IP2-AS1
rs9481167	6	111908882	6.627	1	PSO	TRAF3IP2
rs174378	6	111908892	4.997	0	PSO	TRAF3IP2-AS1
rs10872070	6	111921804	5.818	22	PSO	TRAF3IP2
rs13210247	6	111922720	8.167	6	PSO	TRAF3IP2
rs6568689	6	111942062	6.706	4	PSO	-
rs12526098	6	111942525	5.353	3	PSO	-
rs638173	6	137959044	4.926	3	AGONIST	-
rs514475	6	138038905	4.923	0	AGONIST	-
rs681323	6	138118581	4.861	0	PSO	-
rs55847703	6	138159576	4.981	5	PSO	-
rs674451	6	138216788	6.601	32	PSO	-
rs9497457	6	146462060	4.717	0	ANTAGONIST	GRM1
rs4720243	7	37377580	4.780	4	PSO	ELMO1
rs2700987	7	37386237	5.404	12	PSO	ELMO1
rs2700983	7	37394379	4.890	3	PSO	ELMO1
rs60343814	7	37395851	5.013	2	ANTAGONIST	ELMO1
rs7798970	7	89636224	4.806	7	AGONIST	STEAP2-AS1
rs1554745	8	16330990	4.738	2	ANTAGONIST	MSR1
rs11776712	8	66528677	4.860	0	PSO	ARMC1
rs9406469	9	13855745	4.786	0	AGONIST	-
rs11795343	9	32523737	5.816	3	PSO	DDX58
rs10971028	9	32570122	4.957	1	PSO	NDUFB6

Lead SNV	Chr	Position	T_{max}	# clumped SNVs	Effect	Gene
rs141688385	9	110792282	5.051	5	PSO	-
rs2185975	10	130170176	5.130	11	AE	-
rs4086867	11	26012232	4.726	1	AGONIST	-
rs12295411	11	76275963	5.726	4	AE	-
rs11236797	11	76299649	6.118	22	AE	-
rs72934351	11	76338427	4.795	1	AGONIST	-
rs1131017	12	56435929	5.434	14	PSO	-
rs12580100	12	56439209	5.323	0	PSO	RPS26
rs111492967	12	56731077	6.170	30	PSO	PAN2
rs7317112	13	95923523	4.779	0	AE	ABCC4
chr13:105133101:I	13	105133101	4.731	0	AE	-
rs8016947	14	35832666	5.895	0	PSO	-
rs12883343	14	35852474	4.950	3	PSO	-
rs1528473	15	55386743	5.468	0	AE	-
rs1111186	16	11347048	4.749	0	PSO	SOCS1
chr16:31006289:I	16	31006289	5.261	99	PSO	STX1B
rs9891226	17	12857916	4.958	2	ANTAGONIST	ARHGAP44
rs4795067	17	26106675	5.668	1	PSO	NOS2
rs3794767	17	26124605	4.779	0	PSO	-
rs7237497	18	12777325	4.805	2	AGONIST	-
rs665445	18	51842682	5.270	83	PSO	POLI
rs710845	19	10407169	5.273	1	PSO	ICAM5
rs73510898	19	10416444	5.012	1	PSO	FDX1L
rs2278442	19	10444826	4.792	0	PSO	ICAM3
rs35251378	19	10459969	5.586	4	PSO	TYK2
rs91755	19	10473570	5.674	5	PSO	TYK2
rs10407005	19	10816205	5.347	19	PSO	QTRT1
rs4371271	19	11184293	4.744	0	PSO	-
rs546308	19	49223570	5.407	20	PSO	MAMSTR
chr19:53452475:I	19	53452475	4.826	0	PSO	ZNF816
rs6125833	20	48575471	5.063	8	PSO	KRT18P4
chr20:48577028:I	20	48577028	4.712	0	PSO	KRT18P4
rs6067305	20	48589996	5.845	143	PSO	-
rs6020157	20	48591758	5.278	33	PSO	-
chr20:48617960:I	20	48617960	4.987	3	PSO	-
rs6003725	22	23801430	4.829	3	AGONIST	-

Table A.2.: Distinct loci showing a suggestive association ($T_{max} > 4.7$) by the CCMA method. Chr=chromosome; Position=chromosomal position of the Human Genome Assembly GRCh37/hg19 (Feb. 2009); AE=atopic eczema; PSO=psoriasis, AGONIST=agonistic, ANTAGONIST=antagonistic.

Bibliography

[1] Shanya Sivakumaran, Felix Agakov, Evropi Theodoratou, James G. Prendergast, Lina Zgaga, Teri Manolio, et al. "Abundant pleiotropy in human complex diseases and traits." *Am J Hum Genet* 89(5), 2011, pp. 607–618.

[2] Lam C. Tsoi, Sarah L. Spain, Jo Knight, Eva Ellinghaus, Philip E. Stuart, Francesca Capon, et al. "Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity." *Nat Genet* 44(12), 2012, pp. 1341–1348.

[3] David Ellinghaus, Hansjörg Baurecht, Jorge Esparza-Gordillo, Elke Rodríguez, Anja Matanovic, Ingo Marenholz, et al. "High-density genotyping study identifies four new susceptibility loci for atopic dermatitis." *Nat Genet* 45(7), 2013, pp. 808–812.

[4] Matthias Arnold, Mara L. Hartsperger, Hansjörg Baurecht, Elke Rodríguez, Benedikt Wachinger, Andre Franke, et al. "Network-based SNP meta-analysis identifies joint and disjoint genetic features across common human diseases." *BMC Genomics* 13(490), 2012.

[5] Lucia A. Hindorff, Heather A. Junkins, Jayashri P. Mehta, and Teri A. Manolio. *A Catalog of Published Genome-Wide Association Studies*. www.genome.gov/gwastudies (accessed 9th May 2010). 2008.

[6] David Ellinghaus, Eva Ellinghaus, Rajan P. Nair, Philip E. Stuart, Tõnu Esko, Andres Metspalu, et al. "Combined analysis of genome-wide association studies for Crohn disease and psoriasis identifies seven shared susceptibility loci." *Am J Hum Genet* 90(4), 2012, pp. 636–647.

[7] Samsiddhi Bhattacharjee, Preetha Rajaraman, Kevin B. Jacobs, William A. Wheeler, Beatrice S. Melin, Patricia Hartge, et al. "A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits." *Am J Hum Genet* 90(5), 2012, pp. 821–835.

[8] Mark B. Gerstein, Can Bruce, Joel S. Rozowsky, Deyou Zheng, Jiang Du, Jan O. Korbel, et al. "What is a gene, post-ENCODE? History and updated definition." *Genome Res* 17(6), 2007, pp. 669–681.

[9] Konstantin Strauch. "Kopplungsanalyse bei genetisch komplexen Erkrankungen mit genomischem Imprinting und Zwei-Genort-Krankheitsmodellen (Linkage analyses in complex genetic diseases with genomic imprinting and two-locus-disease-models)". In: *Medizinische Informatik, Biometrie und Epidemiologie*. Ed. by K. Überla, N. Victor, and O. Rienhoff. Vol. 87. München: Urban & Vogel, 2002.

[10] Victor A. McKusick. "Mendelian Inheritance in Man and its online version, OMIM." *Am J Hum Genet* 80(4), 2007, pp. 588–604.

[11] Eric S. Lander and Nicholas J. Schork. "Genetic dissection of complex traits." *Science* 265(5181), 1994, pp. 2037–2048.

[12] David Altshuler, Mark J. Daly, and Eric S. Lander. "Genetic mapping in human disease." *Science* 322(5903), 2008, pp. 881–888.

[13] Eric S. Lander. "The new genomics: global views of biology." *Science* 274(5287), 1996, pp. 536–539.

[14] Neil J. Risch and Kathleen Ries Merikangas. "The future of genetic studies of complex human diseases." *Science* 273(5281), 1996, pp. 1516–1517.

[15] David E. Reich and Eric S. Lander. "On the allelic spectrum of human disease." *Trends Genet* 17(9), 2001, pp. 502–510.

[16] Anthony C. Allison. "Protection afforded by sickle-cell trait against subtertian malareal infection." *Br Med J* 1(4857), 1954, pp. 290–294.

[17] Lars Feuk, Andrew R. Carson, and Stephen W. Scherer. "Structural variation in the human genome." *Nat Rev Genet* 7(2), 2006, pp. 85–97.

[18] Richard S. Spielman, Ralph E. McGinnis, and Warren J. Ewens. "Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)." *Am J Hum Genet* 52(3), 1993, pp. 506–516.

[19] Ravi Sachidanandam, David Weissman, Steven C. Schmidt, Jerzy M. Kakol, Lincoln D. Stein, Gabor Marth, et al. "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms." *Nature* 409(6822), 2001, pp. 928–933.

[20] Goncalo R. Abecasis, Adam Auton, Lisa D. Brooks, Mark A. DePristo, Richard M. Durbin, Robert E. Handsaker, et al. "An integrated map of genetic variation from 1,092 human genomes." *Nature* 491(7422), 2012, pp. 56–65.

[21] Holly K. Tabor, Neil J. Risch, and Richard M. Myers. "Candidate-gene approaches for studying complex genetic traits: practical considerations." *Nat Rev Genet* 3(5), 2002, pp. 391–397.

[22] Anne M. Glazier, Joseph H. Nadeau, and Timothy J. Aitman. "Finding genes that underlie complex traits." *Science* 298(5602), 2002, pp. 2345–2349.

[23] Angela J. Rogers, Benjamin A. Raby, Jessica A. Lasky-Su, Amy Murphy, Ross Lazarus, Barbara J. Klanderman, et al. "Assessing the reproducibility of asthma candidate gene associations, using genome-wide data." *Am J Respir Crit Care Med* 179(12), 2009, pp. 1084–1090.

[24] Kirk E. Lohmueller, Celeste L. Pearce, Malcolm Pike, Eric S. Lander, and Joel N. Hirschhorn. "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease." *Nat Genet* 33(2), 2003, pp. 177–182.

[25] Colin N. A. Palmer, Alan D. Irvine, Ana Terron-Kwiatkowski, Yiwei Zhao, Haihui Liao, Simon P. Lee, et al. "Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis." *Nat Genet* 38(4), 2006, pp. 441–446.

[26] Stephan Weidinger, Thomas Illig, Hansjörg Baurecht, Alan D. Irvine, Elke Rodríguez, Amalia Diaz-Lacava, et al. "Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations." *J Allergy Clin Immunol* 118(1), 2006, pp. 214–219.

[27] Mark J. Daly, John D. Rioux, Stephen F. Schaffner, Thomas J. Hudson, and Eric S. Lander. "High-resolution haplotype structure in the human genome." *Nat Genet* 29(2), 2001, pp. 229–232.

[28] Nila Patil, Anthony J. Berno, David A. Hinds, Wade A. Barrett, Jigna M. Doshi, Coleen R. Hacker, et al. "Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21." *Science* 294(5547), 2001, pp. 1719–1723.

[29] David E. Reich, Stephen F. Schaffner, Mark J. Daly, Gil McVean, James C. Mullikin, John M. Higgins, et al. "Human genome sequence variation and the influence of gene history, mutation and recombination." *Nat Genet* 32(1), 2002, pp. 135–142.

[30] Dana C. Crawford, Tushar Bhangale, Na Li, Garrett Hellenthal, Mark J. Rieder, Deborah A. Nickerson, et al. "Evidence for substantial fine-scale variation in recombination rates across the human genome." *Nat Genet* 36(7), 2004, pp. 700–706.

[31] Sarah A. Tishkoff and Brian C. Verrelli. "Patterns of human genetic diversity: implications for human evolutionary history and disease." *Annu Rev Genomics Hum Genet* 4, 2003, pp. 293–340.

[32] Jakob C. Mueller. "Linkage disequilibrium for different scales and applications." *Brief Bioinform* 5(4), 2004, pp. 355–364.

[33] Stacey B. Gabriel, Stephen F. Schaffner, Huy Nguyen, Jamie M. Moore, Jessica Roy, Brendan Blumenstiel, et al. "The structure of haplotype blocks in the human genome." *Science* 296(5576), 2002, pp. 2225–2229.

[34] Yun Li, Cristen Willer, Serena Sanna, and Gonçalo Abecasis. "Genotype imputation." *Annu Rev Genomics Hum Genet* 10, 2009, pp. 387–406.

[35] Jonathan Marchini, Bryan Howie, Simon Myers, Gil McVean, and Peter Donnelly. "A new multipoint method for genome-wide association studies by imputation of genotypes." *Nat Genet* 39(7), 2007, pp. 906–913.

[36] Olivier Delaneau, Jonathan Marchini, and Jean-François Zagury. "A linear complexity phasing method for thousands of genomes." *Nat Methods* 9(2), 2012, pp. 179–181.

[37] International HapMap Consortium. "A haplotype map of the human genome." *Nature* 437(7063), 2005, pp. 1299–1320.

[38] 1000 Genomes Project Consortium. "A map of human genome variation from population-scale sequencing." *Nature* 467(7319), 2010, pp. 1061–1073.

[39] Illumina. http://www.illumina.com/products/humanomni5-quad_beadchip_kit.ilmn. 2012.

[40] Kelly A Frazer, Sarah S Murray, Nicholas J Schork, and Eric J Topol. "Human genetic variation and its contribution to complex traits." *Nat Rev Genet* 10(4), 2009, pp. 241–251.

[41] Stephan Weidinger, Christian Gieger, Elke Rodríguez, Hansjörg Baurecht, Martin Mempel, Norman Klopp, et al. "Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus." *PLoS Genet* 4(8), 2008, e1000166.

[42] Jorge Esparza-Gordillo, Stephan Weidinger, Regina Fölster-Holst, Anja Bauerfeind, Franz Ruschendorf, Giannino Patone, et al. "A common variant on chromosome 11q13 is associated with atopic dermatitis." *Nat Genet* 41(5), 2009, pp. 596–601.

[43] Keith O'Rourke. "An historical perspective on meta-analysis: dealing quantitatively with varying study results." *J R Soc Med* 100(12), 2007, pp. 579–582.

[44] Karl Pearson. "Report on Certain Enteric Fever Inoculation Statistics." *Br Med J* 2(2288), 1904, pp. 1243–1246.

[45] William G. Cochran. "Problems Arising in the Analysis of a Series of Similar Experiments". *Supplement to the Journal of the Royal Statistical Society* Vol. 4(1), 1937, pp. 102–118.

[46] Gene V. Glass. "Primary, Secondary, and Meta-Analysis of Research". *Educational Researcher* 5(10), 1976, pp. 3–8.

[47] Salim Yusuf, Richard Peto, John Lewis, Rory Collins, and Peter Sleight. "Beta blockade during and after myocardial infarction: an overview of the randomized trials." *Prog Cardiovasc Dis* 27(5), 1985, pp. 335–371.

[48] Michael Borenstein, Larry V Hedges, Julian P T Higgins, and Hannah R Rothstein. "Preface". In: *Introduction to Meta-Analysis*. West Sussex: Wiley & Sons, 2009, pp. xxi–xxviii.

[49] The Cohrane Collaboration. *The Cochrane Organisational Policy Manual Issue*. http://www.cochrane.org/organisational-policy-manual) (accessed 4th July 2016). 2014 [updated 22th December 2014].

[50] Anna-Bettina Haidich. "Meta-analysis in medical research". *Hippokratia* 14, 2010, pp. 29–37.

[51] J. P. T. Higgins and S. Green, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. Available from www.cochrane-handbook.org: The Cochrane Collaboration, 2011.

[52] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G. Altman, and PRISMA Group. "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement." *PLoS Med* 6(7), 2009, e1000097.

[53] Michael Borenstein, Larry V Hedges, Julian P T Higgins, and Hannah R Rothstein. "Effect Sizes Based on Correlations". In: *Introduction to Meta-Analysis*. West Sussex: Wiley & Sons, 2009. Chap. 6, pp. 41–44.

[54] Michael Borenstein, Larry V Hedges, Julian P T Higgins, and Hannah R Rothstein. "Factors that Affect Precision". In: *Introduction to Meta-Analysis*. West Sussex: Wiley & Sons, 2009. Chap. 8, pp. 51–56.

[55] P. Armitage. "Controversies and achievements in clinical trials." *Control Clin Trials* 5(1), 1984, pp. 67–72.

[56] Rebecca DerSimonian and Nan Laird. "Meta-analysis in clinical trials." *Control Clin Trials* 7(3), 1986, pp. 177–188.

[57] B. J. Biggerstaff and R. L. Tweedie. "Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis." *Stat Med* 16(7), 1997, pp. 753–768.

[58] J. Hartung and G. Knapp. "A refined method for the meta-analysis of controlled clinical trials with binary outcome." *Stat Med* 20(24), 2001, pp. 3875–3889.

[59] Kurex Sidik and Jeffrey N. Jonkman. "A simple confidence interval for meta-analysis." *Stat Med* 21(21), 2002, pp. 3153–3159.

[60] C Robert Paule and John Mandel. "Consensus Values and Weighting Factors". *Journal of Research* 87, 1987, pp. 377–385.

[61] R. J. Hardy and S. G. Thompson. "A likelihood approach to meta-analysis with random effects." *Stat Med* 15(6), 1996, pp. 619–629.

[62] D. A. Follmann and M. A. Proschan. "Valid inference in random effects meta-analysis." *Biometrics* 55(3), 1999, pp. 732–737.

[63] Rebecca DerSimonian and Raghu Kacker. "Random-effects model for meta-analysis of clinical trials: an update." *Contemp Clin Trials* 28(2), 2007, pp. 105–114.

[64] Dan Jackson, Jack Bowden, and Rose Baker. "How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts?" *Journal of Statistical Planning and Inference* 140(4), 2010, pp. 961–970.

[65] Evangelos Kontopantelis and David Reeves. "Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study." *Stat Methods Med Res* 21(4), 2012, pp. 409–426.

[66] Evangelos Kontopantelis and David Reeves. "Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: a comparison between DerSimonian-Laird and restricted maximum likelihood." *Stat Methods Med Res* 21(6), 2012, pp. 657–659.

[67] Rebecca DerSimonian and Nan Laird. "Meta-analysis in clinical trials revisited." *Contemp Clin Trials* 45(Pt A), 2015, pp. 139–145.

[68] Buhm Han and Eleazar Eskin. "Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies." *Am J Hum Genet* 88(5), 2011, pp. 586–598.

[69] Joanna IntHout, John P A. Ioannidis, and George F. Borm. "The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method." *BMC Med Res Methodol* 14(25), 2014.

[70] Kurex Sidik and Jeffrey N. Jonkman. "A note on variance estimation in random effects meta-regression." *J Biopharm Stat* 15(5), 2005, pp. 823–838.

[71] Christian Röver, Guido Knapp, and Tim Friede. "Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies." *BMC Med Res Methodol* 15(1), 2015, p. 99.

[72] Michael Borenstein, Larry V Hedges, Julian P T Higgins, and Hannah R Rothstein. "Identifying and Quantifying Heterogeneity". In: *Introduction to Meta-Analysis*. West Sussex: Wiley & Sons, 2009. Chap. 16, pp. 107–126.

[73] Michael Borenstein, Larry V Hedges, Julian P T Higgins, and Hannah R Rothstein. "Meta-Regression". In: *Introduction to Meta-Analysis*. West Sussex: Wiley & Sons, 2009. Chap. 20, pp. 187–204.

[74] Guido Knapp and Joachim Hartung. "Improved tests for a random effects meta-regression with a single covariate." *Stat Med* 22(17), 2003, pp. 2693–2710.

[75] Simon G. Thompson and Julian P T. Higgins. "How should meta-regression analyses be undertaken and interpreted?" *Stat Med* 21(11), 2002, pp. 1559–1573.

[76] Michael Borenstein, Larry V Hedges, Julian P T Higgins, and Hannah R Rothstein. "Subgroup Analysis". In: *Introduction to Meta-Analysis*. West Sussex: Wiley & Sons, 2009. Chap. 19, pp. 149–186.

[77] Michael Borenstein, Larry V Hedges, Julian P T Higgins, and Hannah R Rothstein. "Independent Subgroup within a Study". In: *Introduction to Meta-Analysis*. West Sussex: Wiley & Sons, 2009. Chap. 23, pp. 217–224.

[78] Michael Borenstein, Larry V Hedges, Julian P T Higgins, and Hannah R Rothstein. "Multiple Outcomes or Time-Points within a Study". In: *Introduction to Meta-Analysis*. West Sussex: Wiley & Sons, 2009. Chap. 24, pp. 225–238.

[79] Georgia Salanti, Julian P T. Higgins, A. E. Ades, and John P A. Ioannidis. "Evaluation of networks of randomized trials." *Stat Methods Med Res* 17(3), 2008, pp. 279–301.

[80] Neil Hawkins, David A. Scott, and Beth Woods. "'Arm-based' parameterization for network meta-analysis." *Res Synth Methods*, 2015.

[81] Georgia Salanti and Julian P. T. Higgins. "Meta-analysis of genetic association studies under different inheritance models using data reported as merged genotypes." *Stat Med* 27(5), 2008, pp. 764–777.

[82] Dan Jackson, Martin Law, Jessica K. Barrett, Rebecca Turner, Julian Pt Higgins, Georgia Salanti, et al. "Extending DerSimonian and Laird's methodology to perform network meta-analyses with random inconsistency effects." *Stat Med* 35(6), 2015, pp. 819–39.

[83] Dan Jackson, Richard Riley, and Ian R. White. "Multivariate meta-analysis: potential and promise." *Stat Med* 30(20), 2011, pp. 2481–2498.

[84] Theodore D. Sterling. "Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance–Or Vice Versa". *Journal of the American Statistical Association* 54(285), 1959, pp. 30–34.

[85] R Rosenthal. "The "File drawer problem" and tolerance of null results." *Psychological Bulletin* 86, 1979, pp. 638–641.

[86] Michael Borenstein, Larry V Hedges, Julian P T Higgins, and Hannah R Rothstein. "Publication Bias". In: *Introduction to Meta-Analysis*. West Sussex: Wiley & Sons, 2009. Chap. 30, pp. 277–294.

[87] Jonathan AC Sterne, Matthias Egger, and David Moher. "Addressing reporting bias". In: *Cochrane Handbook for Systematic Reviews of Interventions*. Ed. by Julian PT Higgins and Sally Green. http://handbook.cochrane.org. 2011. Chap. 10.

[88] Matthias Egger, George D. Smith, Martin Schneider, and Christoph Minder. "Bias in meta-analysis detected by a simple, graphical test." *BMJ* 315(7109), 1997, pp. 629–634.

[89] Sue Duval and Richard Tweedie. "A Nonparametric "Trim and Fill" Method of Accounting for Publication Bias in Meta-Analysis". *Journal of the American Statistical Association* 95(449), 2000, pp. 89–98.

[90] S. Duval and R. Tweedie. "Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis." *Biometrics* 56(2), 2000, pp. 455–463.

[91] John P. A. Ioannidis and Thomas A. Trikalinos. "The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey." *Canadian Medical Association Journal* 176(8), 2007, pp. 1091–1096.

[92] Eleftheria Zeggini, Laura J. Scott, Richa Saxena, Benjamin F. Voight, Jonathan L. Marchini, Tianle Hu, et al. "Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes." *Nat Genet* 40(5), 2008, pp. 638–645.

[93] Evangelos Evangelou and John P A. Ioannidis. "Meta-analysis methods for genome-wide association studies and beyond." *Nat Rev Genet* 14(6), 2013, pp. 379–389.

[94] Noah Zaitlen, Bogdan Paşaniuc, Tom Gur, Elad Ziv, and Eran Halperin. "Leveraging genetic variability across populations for the identification of causal variants." *Am J Hum Genet* 86(1), 2010, pp. 23–33.

[95] Andrew P. Morris. "Transethnic meta-analysis of genomewide association studies." *Genet Epidemiol* 35(8), 2011, pp. 809–822.

[96] Hansjörg Baurecht, Alan D. Irvine, Natalija Novak, Thomas Illig, Bettina Bühler, Johannes Ring, et al. "Toward a major risk factor for atopic eczema: meta-analysis of filaggrin polymorphism data." *J Allergy Clin Immunol* 120(6), 2007, pp. 1406–1412.

[97] Elke Rodríguez, Hansjörg Baurecht, Esther Herberich, Stefan Wagenpfeil, Sara J. Brown, Heather J. Cordell, et al. "Meta-analysis of filaggrin polymorphisms in eczema and asthma: robust risk factors in atopic disease." *J Allergy Clin Immunol* 123(6), 2009, 1361–70.e7.

[98] Lavinia Paternoster, Marie Standl, Chih-Mei Chen, Adaikalavan Ramasamy, Klaus Bønnelykke, Liesbeth Duijts, et al. "Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis." *Nat Genet* 44(2), 2011, pp. 187–192.

[99] Lavinia Paternoster, Marie Standl, Johannes Waage, Hansjörg Baurecht, Melanie Hotze, David P. Strachan, et al. "Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis." *Nat Genet* 47(12), 2015, pp. 1449–1456.

[100] Cristen J. Willer, Yun Li, and Gonçalo R. Abecasis. "METAL: fast and efficient meta-analysis of genomewide association scans." *Bioinformatics* 26(17), 2010, pp. 2190–2191.

[101] J. E. Taylor, K. J. Worsley, and F. Gosselin. "Maxima of Discretely Sampled Random Fields, with an Application to 'Bubbles'". *Biometrika* 94, 2007, pp. 1–18.

[102] Hansjörg Baurecht, Melanie Hotze, Elke Rodríguez, Judith Manz, Stephan Weidinger, Heather J. Cordell, et al. "Compare and Contrast Meta Analysis (CCMA): A Method for Identification of Pleiotropic Loci in Genome-Wide Association Studies." *PLoS One* 11(5), 2016, e0154872.

[103] J. L. Fleiss. "The statistical basis of meta-analysis." *Stat Methods Med Res* 2(2), 1993, pp. 121–145.

[104] C. Clopper and E.S. Pearson. "The use of confidence or fiducial limits illustrated in the case of the binomial". *Biometrika* 26, 1934, pp. 404–13.

[105] Warren Ewens and Gregory Grant. *Statisical Methods in Bioinformatics: An Introduction*. Ed. by M Gail, K Krickeberg, and J Samet. 2nd. Statistics for Biology and Health. New York: Springer, 2005, p. 597.

[106] Hansjörg Baurecht, Melanie Hotze, Stephan Brand, Carsten Büning, Paul Cormican, Aiden Corvin, et al. "Genome-wide Comparative Analysis of Atopic Dermatitis and Psoriasis Gives Insight into Opposing Genetic Mechanisms." *Am J Hum Genet* 96(1), 2015, pp. 104–120.

[107] Ludwig Fahrmeir and Alfred Hamerle. "Mehrdimensionale Zufallsvariablen und Verteilungen". In: *Multivariate statistische Verfahren*. Ed. by Ludwig Fahrmeir, Alfred Hamerle, and Gerhard Tutz. Vol. 2. Berlin - New York: de Gruyter, 1996, p. 26.

[108] Evgeny Spodarev. *Stochastik I - Vorlesungsskript*. German. https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/mitarbeiter/spodarev/publications/scripts/Stochastik1_01.pdf (Accessed 5th October 2015). Ulm: University Ulm, 2015.

[109] Zakhar Kabluchko. *Stochastik I (Statistik)*. https://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/lehre/ss13/Stochastik_I/Skript_Stochastik_I.pdf (Accessed 11th July 2016). 2013.

[110] Samsiddhi Bhattacharjee, Nilanjan Chatterjee, and William Wheeler. *ASSET: An R package for subset-based association analysis of heterogeneous traits and subtypes*. R package version 1.4.0. 2013.

[111] R Development Core Team. *R: A Language and Environment for Statistical Computing*. http://www.R-project.org. R Foundation for Statistical Computing. Vienna, Austria, 2010.

[112] Francis Smart. *Simulating Multinomial logit in Stata - Updated*. http://www.econometrics-bysimulation.com/2012/07/simulating-multinomial-logit-in-stata.html (Accessed 23rd June 2015). 2012.

[113] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. 4th ed. New York: Springer, 2002.

[114] Ludwig Fahrmeir and Gerhard Tutz. "Models for Multicategorical Responses: Multivariate Extensions of Generalized Linear Models". In: *Multivariate Statistical Modelling Based on Generalized Linear Models*. 2nd. Springer, 2000. Chap. 3, pp. 71–137.

[115] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, et al. "PLINK: a tool set for whole-genome association and population-based linkage analyses." *Am J Hum Genet* 81(3), 2007, pp. 559–575.

[116] Tomomitsu Hirota, Atsushi Takahashi, Michiaki Kubo, Tatsuhiko Tsunoda, Kaori Tomita, Masafumi Sakashita, et al. "Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population." *Nat Genet* 44(11), 2012, pp. 1222–1226.

[117] Stephan Weidinger and Natalija Novak. "Atopic dermatitis." *Lancet* 387(10023), 2016, pp. 1109–1122.

[118] Chandrashekhar V. Patel, Indhira Handy, Tiffany Goldsmith, and Rekha C. Patel. "PACT, a stress-modulated cellular activator of interferon-induced double-stranded RNA-activated protein kinase, PKR." *J Biol Chem* 275(48), 2000, pp. 37993–8.

[119] Kin-Hang Kok, Pak-Yin Lui, Ming-Him James Ng, Kam-Leung Siu, Shannon Wing Ngor Au, and Dong-Yan Jin. "The double-stranded RNA-binding protein PACT functions as a cellular activator of RIG-I to facilitate innate antiviral response." *Cell Host Microbe* 9(4), 2011, pp. 299–309.

[120] James T. Elder, Allen T. Bruce, Johann E. Gudjonsson, Andrew Johnston, Philip E. Stuart, Trilokraj Tejasvi, et al. "Molecular dissection of psoriasis: integrating genetics and biology." *J Invest Dermatol* 130(5), 2010, pp. 1213–1226.

[121] Julie Henry, Chiung-Yueh Hsu, Marek Haftek, Rachida Nachat, Heleen D. de Koning, Isabelle Gardinal-Galera, et al. "Hornerin is a component of the epidermal cornified cell envelopes." *FASEB J* 25(5), 2011, pp. 1567–76.

[122] Emma Guttman-Yassky, Mayte Suárez-Fariñas, Andrea Chiricozzi, Kristine E. Nograles, Avner Shemer, Judilyn Fuentes-Duculan, et al. "Broad defects in epidermal cornification in atopic dermatitis identified through genomic analysis." *J Allergy Clin Immunol* 124(6), 2009, 1235–1244.e58.

[123] Ulrike Hüffmeier, Heiko Traupe, Vinzenz Oji, Jesús Lascorz, Markward Ständer, Jörg Lohmann, et al. "Loss-of-function variants of the filaggrin gene are not major susceptibility factors for psoriasis vulgaris or psoriatic arthritis in German patients." *J Invest Dermatol* 127(6), 2007, pp. 1367–1370.

[124] Tong Tong Wu, Yi Fang Chen, Trevor Hastie, Eric Sobel, and Kenneth Lange. "Genome-wide association analysis by lasso penalized logistic regression." *Bioinformatics* 25(6), 2009, pp. 714–721.

[125] Stephan Weidinger, Saffron A G. Willis-Owen, Yoichiro Kamatani, Hansjörg Baurecht, Nilesh Morar, Liming Liang, et al. "A genome-wide association study of atopic dermatitis identifies loci with overlapping effects on asthma and psoriasis." *Hum Mol Genet* 22(23), 2013, pp. 4841–4856.

[126] Alexandra Zhernakova, Cleo C. van Diemen, and Cisca Wijmenga. "Detecting shared pathogenesis from the shared genetics of immune-related diseases." *Nat Rev Genet* 10(1), 2009, pp. 43–55.

[127] Emma Guttman-Yassky, Kristine E. Nograles, and James G. Krueger. "Contrasting pathogenesis of atopic dermatitis and psoriasis–part I: clinical and pathologic concepts." *J Allergy Clin Immunol* 127(5), 2011, pp. 1110–1118.

[128] Saffron A G. Willis-Owen, Nilesh Morar, and Charles A. Willis-Owen. "Atopic dermatitis: insights from linkage overlap and disease co-morbidity." *Expert Rev Mol Med* 9(9), 2007, pp. 1–13.

[129] W. O. Cookson, B. Ubhi, R. Lawrence, G. R. Abecasis, A. J. Walley, H. E. Cox, et al. "Genetic linkage of childhood atopic dermatitis to psoriasis susceptibility loci." *Nat Genet* 27(4), 2001, pp. 372–373.

[130] T. Henseler and E. Christophers. "Disease concomitance in psoriasis." *J Am Acad Dermatol* 32(6), 1995, pp. 982–986.

[131] Rajan P. Nair, Kristina Callis Duffin, Cynthia Helms, Jun Ding, Philip E. Stuart, David Goldgar, et al. "Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways." *Nat Genet* 41(2), 2009, pp. 199–204.

[132] Andrew J. Gunderson, Javed Mohammed, Frank J. Horvath, Michael A. Podolsky, Cherie R. Anderson, and Adam B. Glick. "CD8(+) T cells mediate RAS-induced psoriasis-like skin inflammation through IFN-γ." *J Invest Dermatol* 133(4), 2013, pp. 955–963.

[133] Dirkjan Hijnen, Edward F. Knol, Yoony Y. Gent, Barbara Giovannone, Scott J P. Beijn, Thomas S. Kupper, et al. "CD8(+) T cells in the lesional skin of atopic dermatitis and psoriasis patients are an important source of IFN-γ, IL-13, IL-17, and IL-22." *J Invest Dermatol* 133(4), 2013, pp. 973–979.

[134] Ashutosh K. Mangalam, Veena Taneja, and Chella S. David. "HLA class II molecules influence susceptibility versus protection in inflammatory diseases by determining the cytokine profile." *J Immunol* 190(2), 2013, pp. 513–518.

[135] H. C. Williams, P. G. Burney, A. C. Pembroke, and R. J. Hay. "The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. III. Independent hospital validation." *Br J Dermatol* 131(3), 1994, pp. 406–416.

[136] Michael Krawczak, Susanna Nikolaus, Huberta von Eberstein, Peter J P. Croucher, Nour Eddine El Mokhtari, and Stefan Schreiber. "PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships." *Community Genet* 9(1), 2006, pp. 55–61.

[137] H-E. Wichmann, C. Gieger, T. Illig, and MONICA/KORA Study Group. "KORAgen–resource for population genetics, controls and a broad spectrum of disease phenotypes." *Gesundheitswesen* 67 Suppl 1, 2005, S26–S30.

[138] S. K. Weiland, E. von Mutius, T. Hirsch, H. Duhme, C. Fritzsch, B. Werner, et al. "Prevalence of respiratory and atopic disorders among children in the East and West of Germany five years after unification." *Eur Respir J* 14(4), 1999, pp. 862–870.

[139] Michael C. O'Donovan, Nicholas Craddock, Nadine Norton, Hywel Williams, Timothy Peirce, Valentina Moskvina, et al. "Identification of loci associated with schizophrenia by genome-wide association and follow-up." *Nat Genet* 40(9), 2008, pp. 1053–1055.

[140] Amy Strange, Francesca Capon, Chris C A. Spencer, Jo Knight, Michael E. Weale, Michael H. Allen, et al. "A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1." *Nat Genet* 42(11), 2010, pp. 985–990.

[141] David G. Clayton, Neil M. Walker, Deborah J. Smyth, Rebecca Pask, Jason D. Cooper, Lisa M. Maier, et al. "Population structure, differential bias and genomic control in a large-scale, case-control association study." *Nat Genet* 37(11), 2005, pp. 1243–1246.

Eidesstattliche Versicherung

(**Siehe Promotionsordnung vom 12.07.2011, §8, Abs. 2 Pkt. 5**)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 13. August 2016

Hansjörg Baurecht
(Unterschrift)

Erklärung

(**Siehe Promotionsordnung vom 12.07.2011, §8, Abs. 2 Pkt. 7**)

Hiermit erkläre ich, dass die Dissertation mit dem Titel "Compare and Contrast Meta Analysis (CCMA): An Application for Genomewide Association Studies" keiner anderen Prüfungskommission vorgelegt worden ist.

München, den 13. August 2016

Hansjörg Baurecht
(Unterschrift)

Erklärung

(**Siehe Promotionsordnung vom 12.07.2011, §8, Abs. 2 Pkt. 9**)

Hiermit erkläre ich, dass ich mich **nicht** anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

München, den 13. August 2016

Hansjörg Baurecht
(Unterschrift)

Acknowledgement

I would like to thank all the people who supported me in completing this thesis.

Especially I have to thank my Ph.D. advisor Prof. Dr. Konstantin Strauch for his excellent supervision, his very helpful and thorough feedback and giving me free rein of choosing the subject,

Prof. Dr. Thomas Augustin who gave me the opportunity of doing a Ph.D. at the Fakultät für Mathematik, Informatik und Statistik der Ludwig-Maximilians-Universität München and the evaluation of this thesis as second examiner,

my colleague Melanie Hotze, Prof. Dr. Heather Cordell and Dr. Sara Brown for their contribution and enthusiasm pushing the joint project of comparing & contrasting psoriasis and atopic eczema to publication, of which the developed statistical method was an essential component,

the head of my work group and friend Prof. Dr. Stephan Weidinger for his constant support, employment and promotion as a researcher in the field of dermatogenetics over the last years,

my sister in law Naa Anorkor Baurecht-Abbey and my colleague and friend Dr. Elke Rodríguez for their steady encouragement and proof-reading,

my very best friends Martin and Stephan for being so proud of me and pushing me to the end of this thesis,

my parents Annemarie and Martin Baurecht and my sister and brother Andrea and Alexander for their constant support in all the good and bad times.