UNDERSTANDING HETEROGENEITY AND INTERACTION IN THE CONTEXT OF
WHOLE GENOME GENETIC ANALYSIS

by

Sriram Balu

M.D., Kharkov State Medical University, 1996

Submitted to the Graduate Faculty of
Department of Human Genetics
Graduate School of Public Health in partial fulfillment
of the requirements for the degree of
Master of Science

University of Pittsburgh
2007
This thesis was presented

by

Sriram Balu

It was defended on

January 30, 2007

and approved by

Thesis Advisor:
Dr. Michael M. Barmada, Ph.D.
Associate Professor, Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Dr. Eleanor Feingold, Ph.D.
Associate Professor, Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Dr. John W. Wilson, Ph.D.
Assistant Professor, Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh
Interactions and heterogeneity play a vital role in the miscommunication between genotype and phenotype in complex diseases. Detection of genes that influence the risk of common, complex disorders involves many statistical and computational challenges. This led us to investigate and compare the common methods of linkage analysis in complex diseases. We applied various methods of linkage analysis on the simulated dataset from the Genetic Analysis Workshop (GAW) 14. As the disease modeled in this dataset resembled a qualitative disorder, we employed methods such as nonparametric linkage scans, association studies of susceptibility regions, and conditional studies (conditioning on a previously identified susceptibility locus).

The goal of this project was to study the efficiencies and inadequacies of various methods in detecting interactions and heterogeneity in the simulated dataset.

The methods used on this dataset showed very low percentage in the detection of interactions. We attribute this unsatisfactory performance of these methods mostly to the low
prevalence of interactions in the imaginary populations studied. We also propose various ways of
improving the power in these analyses like considering haplotype studies instead of targeting
single markers and increasing the range of the flanking markers around regions of high LOD
scores.

**Public Health Importance:** Understanding the complexities involved in the genetics of diseases
will provide new insight for disease prevention and health promotion. For over twenty years,
public health agencies have focused more and more on newborn screening programs to detect
and prevent rare genetic disorders. But common complex disorders pose a bigger problem
because of their unique characteristics like heterogeneity, gene-gene interactions, multiple
susceptible loci, incomplete penetrance, phenocopy and presence of environmental risk factors.
By comparing common methods of linkage analysis in complex disorders in the simulated
dataset of Genetic Analysis Workshop (GAW) 14, our study aims to come up with a better
understanding of how heterogeneity and interaction work in the context of a whole genome
genetic analysis. It is also expected to lay a foundation on which future public health researchers
will be able to expand on our work.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS........................................................................................................................................... ix
LIST OF ABBREVIATIONS...................................................................................................................................... xi

1.0 INTRODUCTION ......................................................................................................................................... 1
2.0 OBJECTIVES .............................................................................................................................................. 2
3.0 LITERATURE REVIEW .............................................................................................................................. 3
  3.1 SIMPLE APPROACH TO COMPLEX TRAITS ................................................................................... 3
  3.1.1 Incomplete Penetrance and Phenocopy ...................................................................................... 4
  3.1.2 Genetic Heterogeneity ................................................................................................................. 5
  3.1.3 Polygenic inheritance ..................................................................................................................... 6
  3.1.4 Gene-gene interaction .................................................................................................................. 6
  3.1.5 Gene-environment interaction ................................................................................................... 8

3.2 STUDYING COMPLEX TRAITS – APPROACHES AND LIMITATIONS ...................................... 10
  3.2.1 Linkage-based approaches ........................................................................................................... 10
    3.2.1.1 Parametric linkage analysis ................................................................................................. 11
    3.2.1.2 Non-parametric linkage analysis ......................................................................................... 12
  3.2.2 Linkage disequilibrium (LD) based approaches ....................................................................... 13
    3.2.2.1 Case-control association studies ......................................................................................... 14
    3.2.2.2 Family-based association tests............................................................................................ 15
    3.2.2.3 Haplotype-based approaches .............................................................................................. 16

3.3 INTRODUCTION TO GENETIC ANALYSIS WORKSHOP .................................................... 17

4.0 METHODS .................................................................................................................................................. 22
  4.1 OVERVIEW ............................................................................................................................................ 22
  4.2 NONPARAMETRIC LINKAGE ANALYSIS ....................................................................................... 23
  4.3 ANALYSES ON THE PURCHASED DATA ..................................................................................... 24
    4.3.1 Linkage analysis and TDT Analysis ......................................................................................... 24
    4.3.2 Interaction Analyses ............................................................................................................... 24
4.3.2.1 Stratification ................................................................. 25
4.3.2.2 Weighted analysis ......................................................... 27
4.3.2.3 Redefinition ................................................................. 29
4.3.2.4 Logistic Regression ....................................................... 30
5.0 RESULTS .................................................................................. 34
5.1 INITIAL LINKAGE ANALYSIS .................................................. 34
5.2 ANALYSES ON THE PURCHASED DATA ............................... 35
  5.2.1 Linkage Analysis and Association Analysis ....................... 35
  5.2.2 Interaction Analysis ............................................................ 37
    5.2.2.1 Stratification .............................................................. 37
    5.2.2.2 Weighted Analysis ..................................................... 44
    5.2.2.3 Redefinition .............................................................. 50
    5.2.2.4 Logistic Regression .................................................... 55
6.0 DISCUSSION ............................................................................ 57
7.0 CONCLUSION ........................................................................... 60
8.0 FUTURE STUDIES ..................................................................... 63
APPENDIX A. SHELL SCRIPT FOR NON-PARAMETRIC LINKAGE ANALYSIS .... 64
APPENDIX B. SHELL SCRIPT FOR LOGISTIC REGRESSION ANALYSIS .......... 243
BIBLIOGRAPHY .............................................................................. 250
LIST OF TABLES

Table 1. Structure of the GAW populations. ................................................................. 18
Table 2. Ascertainment scheme used to construct the datasets. ................................. 19
Table 3. Disease model summary. ................................................................................ 20
Table 4. Location of disease related loci. ..................................................................... 21
Table 5. Results from genome scan with microsatellite and snp genome scan markers. ....34
Table 6. Overtransmitted alleles for regions with highest linkage signals on the genome....36
Table 7. Nonparametric analyses in the stratified groups – replicate 1 .........................38
Table 8. Nonparametric analyses in the stratified groups – replicate 2 .........................39
Table 9. Nonparametric analyses in the stratified groups – combined replicates 1 and 2. 40
Table 10. Nonparametric analyses in the stratified groups – combined replicates 1 through 5. 41
Table 11. Simulations on stratified groups – replicate 1. ..............................................42
Table 12. Simulations on stratified groups – replicate 2. ..............................................43
Table 13. Simulations on stratified groups – combined replicates 1 and 2. ....................43
Table 14. Simulations on stratified groups – combined replicates 1 through 5. ............44
Table 15. Nonparametric analyses and simulations in the weighted groups – replicate 1. 46
Table 16. Nonparametric analyses and simulations in the weighted groups – replicate 2. 47
Table 17. Nonparametric analyses in the weighted groups – combined replicates 1 and 2. 47
Table 18. Nonparametric analyses in the weighted groups – combined replicates 1 through 5. 49
Table 19. Nonparametric analyses in the redefined groups – replicate 1. ......................51
Table 20. Nonparametric analyses in the redefined groups – replicate 2. ......................52
Table 21. Nonparametric analyses in the redefined groups – combined replicates 1 and 2. 53
Table 22. Nonparametric analyses in the redefined groups – combined replicates 1 through 5. 54
Table 23. Logistic regression analyses using the program lrmodel. ..............................56
Table 24. Comparison of results of conditional analyses with the GAW ‘answers’. ..........61
Table 25. Prevalence of interactions. ..........................................................................62
LIST OF FIGURES

Figure 1. Incomplete penetrance ..................................................................................................... 4
Figure 2. Linkage analysis ............................................................................................................. 10
Figure 3. Affected sib-pair (ASP) analysis ................................................................................... 12
Figure 4. Case-control association analysis .................................................................................. 14
Figure 5. Family-based association analysis ............................................................................... 15
Figure 6. Comparison of lod scores from npl analyses on purchased data ......................... 35
ACKNOWLEDGEMENTS

It is my primary duty to express my indebtedness to my guide and mentor, Dr. Michael Barmada, for the unwavering support I have been receiving from him right from the day I started my work under him. Sparing indeed with his time and erudition, yet, he was unsparing in his demand for excellence and perfection in the research I had taken up. Personifying ‘caring and sharing’ in the true sense of the term, Dr. Michael Barmada understood my problems and offered solutions forthwith.

I owe an ocean of gratitude to Dr. Robert Ferrell, my academic advisor, whose professional knowledge can be matched only by his own. An epitome of meticulousness and commitment, his suggestions and academic insights have gone a long way in shaping my research into its present form.

I cannot adequately thank the committee members, Dr. Daniel Weeks, Dr. Eleanor Feingold, Dr. Candace Kammerer, Dr. David Finegold and Dr. John Wilson, for the readiness and willingness with which they helped me out whenever I approached them. I thank them individually for all the timely tips and for all the succor and support I had received from them.

I am grateful to Dr. Susanne Gollin, Dr. Urvashi Surti and Dr. Sofia Shekhter-Levin for the motivation I had received from them and for the kindness shown to me from time to time.
My thanks are due to Michaele Armstrong, my batch mate, for the social, moral and academic support she had been offering me with an ever-smiling face.

I thank every member of my family in general and Chitra Chauhan in particular for being patiently behind me in my present endeavor and for generously pardoning me for my having overburdening them.

I thank one and all with whom I have been involved in this project.
LIST OF ABBREVIATIONS

chr: chromosome
htSNP: haplotype tag single nucleotide polymorphism
jck: juvenile cystic kidney
p-value: probability value
snp: single nucleotide polymorphism
AFBAC: affected family-based controls
AIC: Akaike’s information criterion
APM: affected pedigree member
APOE: apolipoprotein E
ASP: affected sib-pair
BRCA1: breast cancer gene 1
BRCA2: breast cancer gene 2
CARD: caspase recruitment domain
CD: Crohn’s disease
CYP: cytochrome P-450 enzyme
DNA: deoxyribonucleic acid
GAW: genetic analysis workshop
GLM: generalized linear model
HRR: haplotype relative risk
IBD: identical-by-descent
IBD: inflammatory bowel disease
IBS: identical-by-state
LD: linkage disequilibrium
LOD: logarithm of odds
LRS: likelihood ratio statistic
NIDDM: non-insulin-dependent diabetes mellitus
NOD: nucleotide-binding oligomerization domain
NPL: non-parametric linkage
OMIM: online Mendelian inheritance in man
PDT: pedigree disequilibrium test
PKD: polycystic kidney disease
PKU: phenylketonuria
TDT: transmission disequilibrium test
1.0 INTRODUCTION

The development of any complex disease is an active process that is influenced by a system of genes as well as by environmental factors. It is a statistical and computational challenge to identify and characterize genes that influence the risk of common, complex multifactorial disease. The presence of heterogeneity and interactions with other genes and environmental factors makes the task even harder.

Genetic linkage mapping is a systematic, genome-wide approach to the study of complex diseases. Genetic linkage studies of complex diseases have certainly identified susceptible chromosomal regions and provide a reliable basis for additional linkage and association studies. At least a few loci are supported by several linkage or association studies.

We set out to study common methods of linkage analysis in complex disorders in the simulated data set of Genetic Analysis Workshop (GAW) 14. As this set of simulated data mimicked a qualitative disorder, we used methods which are employed in other complex qualitative disorders, such as nonparametric linkage scans, association studies of susceptibility regions, and conditional studies (conditioning on a previously identified susceptibility locus).
2.0 OBJECTIVES

The primary objective of this study is to come up with a better understanding of how heterogeneity and interaction look from the point of view of a whole genome scan – is the evidence of linkage inflated at both loci or only one? Does conditioning on one locus find the other one if there is an interaction or only when there is heterogeneity?

To achieve these objectives we propose the following aims:

1. to devise an effective scheme to study methods of linkage analysis in the simulated data set of GAW14
2. to perform conditional analyses by applying stratification methods
3. to perform conditional analyses by applying redefinition methods
4. to perform conditional analyses by applying weighted analyses
5. to perform conditional analyses by applying logistic regression methods
3.0 LITERATURE REVIEW

3.1 SIMPLE APPROACH TO COMPLEX TRAITS

Traits that are influenced by multiple loci (genes) are called complex traits. There is no universal definition of complex trait in the literature. In a broad sense, it refers to any phenotype that does not display classic Mendelian recessive or dominant inheritance characteristic of a single gene locus. However, compartmentalizing genetic disorders into simple/monogenic/Mendelian and complex/multifactorial/non-Mendelian might be an oversimplification. Many diseases that were initially thought to be monogenic are turning out to be either caused or modulated by the action of a small number of loci. Even the most classic ‘monogenic disorders’, like Phenylketonuria (PKU) and sickle cell anemia, are now being considered as part of the spectrum of complex disorders. These disorders are described as ‘oligogenic’ disorders—a wide range of phenotypes that are neither monogenic nor complex. Nonetheless, the etiology of most of these disorders is due to the presence of a major gene of effect with a clear inheritance pattern thereby making it relatively much easier to identify the primary genetic cause. This is evident in the OMIM database (http://www.ncbi.nlm.nih.gov/Omim/) with entry of ~2000 human diseases or traits with known molecular basis, a fairly large proportion of which are Mendelian disorders. The success attained
for the Mendelian disorders could be quite inspiring for gene-mapping efforts for complex disorders. However, there are several characteristics of the complex disorders that provide daunting challenge to such efforts.

### 3.1.1 Incomplete Penetrance and Phenocopy

Incomplete penetrance implies that individuals inheriting a predisposition allele may not always manifest the disease. Thus a given genotype may affect the probability of the disease but may not completely determine the outcome. Incomplete penetrance manifests itself as the disease occasionally skipping generations (Figure 1).

![Figure 1. Incomplete penetrance. A disease skipping generations is a hallmark of incomplete penetrance](image)

Mutations in *BRCA1* and *BRCA2* genes in the case of rare forms of breast cancer, and in *APOE* gene in the case of Alzheimer’s disease are examples of incomplete penetrance.
Incomplete penetrance may be caused by interaction with other loci or environment or it could be purely a chance event. On the other hand, there may be some individuals who are affected with the same disorder but due to purely non-genetic (e.g. environmental) reasons. Such individuals (phenocopies) are clinically indistinguishable from individuals harboring the predisposition allele. Both of these phenomena hamper the gene-mapping process as the predisposition allele may be present in some unaffected individuals (incomplete penetrance) or absent in some affected individuals (phenocopy).

3.1.2 Genetic Heterogeneity

Genetic heterogeneity is a phenomenon whereby distinct mutations at the same locus (allelic heterogeneity) or at different loci (non-allelic heterogeneity) can cause the same indistinguishable phenotype. It is the non-allelic heterogeneity which is the hallmark of complex disorders. This is generally the case when different genes in the same biochemical pathway harbor the predisposition allele. For example recessive retinitis pigmentosa can be caused by at least 40 genes\textsuperscript{95}, whereas congenital hearing loss can be observed to be the result of mutations in at least 70 independent loci\textsuperscript{88}. Such a condition poses problems to medical geneticists who cannot distinguish between patients suffering from the same disease for different genetic reasons. Non-allelic heterogeneity hampers genetic mapping efforts in many ways. It often results in a single gene accounting for a small proportion of segregating families. In such situations very large families are required to obtain robust linkage evidence. Moreover, in the case of common complex disorders, in a large pedigree, chances are high that individuals carrying different predisposition genes marry into the pedigree. Then the investigator may in fact be looking at linkage of two more loci at the same time, making the linkage analysis much more difficult.\textsuperscript{68}
Since different chromosomal regions may be involved with the disease in different families, it has been one of the major reasons for non-replication of initial gene-mapping findings’ localization and conflicting position estimates of disease loci.\textsuperscript{81}

### 3.1.3 Polygenic inheritance

In the case of many complex disorders, genetic heterogeneity and polygenic inheritance go together, making the gene-mapping even more complicated. Polygenic inheritance implies that multiple genetic variants, within the same or different genes, combine to affect liability for many common diseases. The variants may interact among themselves and with environmental factors (gene-gene and gene-environment interactions respectively). Polygenic traits may be classified as discrete traits or quantitative traits. Discrete traits may represent a threshold effect, produced whenever an underlying quantitative variable, influenced by multiple genes, exceeds a critical threshold, or a pure synthetic effect, requiring the simultaneous and joint action of each of several mutations.\textsuperscript{50} Polygenic inheritance obscures genetic mapping efforts because no single locus is individually necessary and sufficient to produce a discrete trait or a high value of a quantitative trait.

### 3.1.4 Gene-gene interaction

Gene-gene interaction is thought to be an important component of the genetic architecture of complex disorders. The involvement of biomolecular interactions in almost every biological process such as gene regulation and metabolism suggests that relationship between DNA sequence variations and clinical endpoints is likely to involve gene-gene interactions.
Epistasis makes the gene-mapping efforts even more daunting as it causes the alteration of the effect of one locus by effects at another locus. In such scenario the power to detect the first locus may get reduced and, in addition, elucidation of the joint effects of the two loci may get hindered by their interaction. If more than two loci are involved, which is the case in most of the complex disorders, the situation is further complicated by the possibility of complex multiway interactions among some or all of the contributing loci.

The following is a summary of the difficulties in the statistical treatment of gene-gene interaction effects³³:

**Sample size and power:** Complex traits are influenced by many genetic and non-genetic factors; to detect the compelling and robust evidence of interaction between them a very large sample size is required.

**Modeling interactions:** Genes can interact in a variety of ways; making assumptions about the specific way in which the interaction manifests itself can be problematic from an analytical point of view, since a large number of parameters might need to be estimated.

**Multiple comparisons:** If there is no *a priori* knowledge about anticipation, then testing for it would require examination of all possible combinations of the variants used for the study. This number can be as high as tens of thousands for two-locus combinations and may be countless for three or more locus combinations in case of genome-wide scans. Things would further complicate if different assumptions are made for each combination of loci. Multiple testing in such cases raises doubts about false-positive findings.
Biological significance: The detection of statistical interaction does not always imply true gene-gene interaction at the biological level. This is particularly the case in genome-wide scans because the use of marker locus genotypes to draw inferences about putative trait loci could be marred by a number of biological factors which might lead to erroneous inferences about interaction.

3.1.5 Gene-environment interaction

The same genotype at a given trait locus may have different effects on the phenotype under different environmental conditions such as the lifestyle, food habits, surrounding environment etc – a phenomenon known as gene-environment interaction. For example, increased intake of fat and calories and reduced physical activity can lead to obesity, diabetes, asthma, and cardiovascular diseases. However, genetic studies generally ignore the environmental effects and their possible interaction with genes. For example, even though the \( \varepsilon_4 \) allele of APOE (apolipoprotein E) gene is strongly associated with the common forms of Alzheimer’s disease, individuals eating fish at least once a week have been shown to have a 60% lower risk for the disease compared with individuals who never or rarely eat fish\(^6\). However, how the interplay between fish consumption and APOE variation affects the risk of this disease is not clear. Similar effect of environment on the risk of a complex disorder has been demonstrated in the case of schizophrenia, wherein a higher risk of developing this disorder has been shown in children born in winter or spring. Prenatal exposure to viral infections, more prevalent during these seasons, is considered as a possible trigger for the causation of this
effect.\textsuperscript{15} Thus, exposure to infectious agents is gaining importance as a likely interacting environmental factor in influencing the risk for complex disorders such as diabetes, heart diseases and obesity.\textsuperscript{14}
3.2 STUDYING COMPLEX TRAITS – APPROACHES AND LIMITATIONS

There are myriad methods and approaches available in today’s fast-growing field of genetic epidemiology. These can be generally classified as either linkage based or linkage disequilibrium (LD) based approaches. The following are examples of such methods and strategies for studying complex traits.

3.2.1 Linkage-based approaches

Linkage analysis examines the co-segregation of a marker locus with a disease gene locus (i.e. a disease phenotype), with no \textit{a priori} knowledge about the physical position of the disease gene (Figure 2). If the correlation between genotype and phenotype is strong then, when the segregation is evaluated statistically, the probability of finding the true disease gene location is high. However, in cases where some genes have only a minor to moderate effect on the phenotype, as in diseases with multiple disease-causing genes, the probability of finding truly linked loci decreases.

\textbf{Figure 2. Linkage analysis.} Linkage analysis involves studying the co-segregation of genetic markers with the disease in large multi-generational families with multiply affected individuals.
Linkage analysis can be divided into two main strategies. I) parametric (or model-based) linkage analysis and II) non-parametric (or model-free) linkage analysis.

### 3.2.1.1 Parametric linkage analysis

Parametric linkage analysis uses multi-case families and extended pedigrees. This method involves calculating a logarithm of odds (LOD) score. The LOD score is computed by first calculating the maximum likelihood estimate - the ratio of a marker being linked or not linked to a disease locus under a specific assumption. To conclude whether apparent linkage is “real,” the concept of “genome-wide significance” has been developed - the probability threshold that declares linkage after testing many DNA markers used in a genome scan. Lander and Kruglyak\(^5\) suggested 3 levels of genome-wide significance: suggestive linkage, significant linkage and confirmed linkage, though it is suggested that confirmed linkage only occurs when the results are replicated in an independent study sample.

Multipoint linkage analysis is often performed to increase the power of a linkage study. In such type of linkage analysis, several markers known to be located in proximity of each other are analyzed together. If there is true linkage, all markers adjacent to the disease locus are expected to be linked.

Although parametric linkage analysis has the highest power for Mendelian disorders and efficient genome scans can be performed efficiently using this method, it is of limited value for mapping genes for heterogeneous complex disorders. The main drawback of this method is the need to specify the mode of transmission of the disease allele (recessive or dominant), the disease allele frequency in the population and the penetrance value, which are not known in case of most of the complex disorders. However, to get around this, when the data are analyzed several times under different models, the highest LOD score is close to the true mode of
inheritance. Moreover, in the case of many complex disorders, particularly for those with late age at onset (for example Alzheimer’s disease), multiply affected large multigenerational families are difficult to collect.

3.2.1.2 Non-parametric linkage analysis

Non-parametric methods avoid problems of selecting a very specific model while still using some of the power of linkage analysis. In these methods affected sib pairs or other pairs of affected relatives are studied and theoretically no assumptions about the disease model are needed to be made. Methods applying this approach are the affected sib-pair method (ASP) using sib-pairs or nuclear families (Figure 3), and the affected pedigree member method (APM) using extended families.

Figure 3. Affected sib-pair (ASP) analysis. ASP analysis involves testing whether affected sib-pairs inherit a region identical-by-decent (IBD) more often than expected under random Mendelian segregation.

The underlying principal behind the non-parametric linkage analysis methods can be explained as follows. Two chromosomal regions are said to be identical-by-descent (IBD) if they
descend from the same ancestral chromosomes. Regions can also be identical-by-state (IBS) if they share the same alleles, but the origin of the region is unknown and therefore the sharing could simply be due to chance. If there is a susceptibility gene located somewhere in the genome and shared by affected individuals IBD, markers physically close to this region will be transmitted along with the disease allele. Furthermore, if a region is shared among affected individuals from the same family more frequently than is expected by random segregation, it may harbor the disease gene. Although, knowing the parental genotypes is not absolutely essential, parental genotypes make such type of analysis more powerful by increasing the certainty about the IBD.

The non-parametric linkage analysis is a method of choice as a first approach for identifying linkage in complex disorders, for example Alzheimer’s disease and alcoholism. However, since the microsatellite markers used for genome-wide analysis are widely spaced, the linked chromosomal region may contain hundreds of genes and so to narrow down the list of such positional candidate genes, additional analyses are required.

### 3.2.2 Linkage disequilibrium (LD) based approaches

LD mapping can be used at the genome-wide level, as a complementary strategy for identifying disease genes within a defined candidate region identified through linkage analysis as well as a candidate gene approach. This approach is based on the fact that alleles at neighboring loci tend to segregate together. If two loci are inherited together more often than would be expected by independent segregation, the two loci are said to be in linkage disequilibrium. It has been suggested that association studies are more powerful than linkage analysis when searching for susceptibility genes for complex disorders. LD mapping can be a useful tool when
performing fine mapping of regions identified by traditional linkage analysis or in the identification of genes of minor effect in complex disorders.

This approach is applied at the population level to perform case-control association analysis, and at the nuclear family level to perform transmission disequilibrium test (TDT) or haplotype relative risk (HRR) analysis. Both case-control and family-based association tests can be performed using single markers or haplotypes.

3.2.2.1 Case-control association studies

The case-control association studies consider specific markers and the allele frequencies between patients and controls are compared (Figure 4). A positive association finding for a specific marker could occur if it was very close to the true disease locus.

Though this approach is considered to be statistically very powerful, case-control association study can generate false positives as a result of population stratification. This problem can be addressed by the use of family-based association methods (discussed below).

![Figure 4](from Burmeister, 1999). Marker allele frequencies are statistically compared between cohorts of unrelated patients and ethnically matched unrelated controls.
3.2.2.2 Family-based association tests

Transmission disequilibrium test (TDT) and haplotype relative risk (HRR) test analyze for preferential transmission of specific alleles (Figure 5). If an association between a genetic marker and a disease locus exists, transmission of the marker locus from parents to an affected offspring would deviate from the expected 0.5 value predicted by Mendelian inheritance. In TDT, only heterozygous parents are considered, while in the HRR test, both homozygous and heterozygous parents are included.

**Figure 5. Family-based association analysis (from Burmeister, 1999).**

A. Family based association analyses start with ascertaining the affected individuals and their parents. B. Haplotype relative risk (HRR) method. The allele that is not present in the patient is used as control. C. Transmission disequilibrium test (TDT). The preferential transmission of an allele from a heterozygous parent to the affected offspring is tested.
Other transmission test methods include pedigree disequilibrium test (PDT),\textsuperscript{56} which incorporates information from all members of a pedigree with genotypic and phenotypic data, and affected family-based controls (AFBAC) test, where both simplex and multiplex families can be ascertained.\textsuperscript{91} The AFBAC test uses family data to estimate control marker frequencies assuming random mating in the population. It uses the parental marker alleles not transmitted to an affected child, or never transmitted to an affected sib-pair, as the control population. As a result, association due to ethnic mismatching between patients and controls is avoided, which might cause biased results in traditional case-control studies with unrelated individuals.

### 3.2.2.3 Haplotype-based approaches

Studying the disease-associated mutations in conjunction with the ancestral haplotypes from which they arose can be statistically powerful methods to apply in association studies of human diseases. Haplotypes created from multiple marker sites can be valuable due to their higher information value relative to single SNPs. The use of haplotype-based approaches has contributed to the identification of genes for both Mendelian disorders\textsuperscript{33} and for diseases with a more complex inheritance.\textsuperscript{12, 41} Individual haplotypes that contain a mutation will be inherited along with the chromosomal region immediately surrounding that mutation. The size of the ancestral segment shared by the haplotypes is affected by the age of the mutation. Alleles carrying a younger mutation tend to share larger segments than older ones. Studies of haplotype block structure show that only a few haplotypes (approximately 3-5) will constitute the majority (around 90\%) of all chromosomes present in a population.\textsuperscript{25} Most of these frequently found haplotypes could be distinguished by only a few “key” SNPs, also called the haplotype tag SNPs (htSNP).\textsuperscript{12, 42, 67} This suggests that haplotype analysis will reinforce LD mapping by significantly reducing the required number of genotypes, making this a more cost-effective approach.\textsuperscript{42}
3.3 INTRODUCTION TO GENETIC ANALYSIS WORKSHOP

The Genetic Analysis Workshops (GAWs) were started in 1982. The main objective of the workshops is to bring together investigators from different parts of the world to put their ideas of research to a single data set. Such an effort was expected to determine the numerical accuracy of the algorithms, to examine the robustness of the methodologies and to compare the range of conclusions from such a data set. Complex traits have been one of the areas of focus of these workshops.

Participants of these workshops are encouraged to interact and cooperate in discussing the various issues related to the data set they worked on, to compare results and interpretations, thereby throwing light on unresolved problems in genetic analysis. These workshops are an excellent place for both beginners and experienced personnel in the research field to get hands on experience in genetic analysis.

Simulated Data from Genetic Analysis Workshop (GAW) 14

The simulated data for GAW 14 includes an imaginary disease that is thought to be genetic in origin with some environmental factors.

The main characteristics of the genetic model in the simulated dataset that led us to investigate it further are:

1. interaction and
2. heterogeneity.

Four populations were generated (Table 1), each with the same underlying population parameters. Three of the populations had nuclear families. One population had large extended
families. The simulated data set was structured in such a manner that it was possible for participants to work on “fine” mapping and to potentially detect association of the disease due to linkage disequilibrium.

Table 1. Structure of the GAW populations.

<table>
<thead>
<tr>
<th>Population</th>
<th>Number of pedigrees</th>
<th>Nature of pedigrees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aipotu (AI)</td>
<td>100</td>
<td>Nuclear families</td>
</tr>
<tr>
<td>Danacaa (DA)</td>
<td>100</td>
<td>Nuclear families</td>
</tr>
<tr>
<td>Karangar (KA)</td>
<td>100</td>
<td>Nuclear families</td>
</tr>
<tr>
<td>New York City (NYC)</td>
<td>50</td>
<td>Extended families</td>
</tr>
</tbody>
</table>

The fine-mapping data were distributed from a specially-designed website that limited the number of SNPs that could be “bought” using an imaginary fund to make it look like a real situation in the management of research funds. The chromosomal regions were packaged into groups of 20 contiguous SNPs (“packets”) and participants were allowed to download a maximum of 20 packets. Each packet contained files for all replicates from all the populations studied.

In order to simulate linkage disequilibrium, an area of the genome was used in which there was some LD and two-SNP haplotype probabilities from that region were used to simulate the data.
Collection of Data

The simulated data include data collections from four different (imaginary) groups of investigators from the (simulated) countries of Aipotu (AI), Danacaa (DA), Karangar (KA) and New York City (NYC) (Table 2).

<table>
<thead>
<tr>
<th>Population</th>
<th>Affection status based on phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Affected if either P1, P2 or P3 is present</td>
</tr>
<tr>
<td>DA</td>
<td>Affected only if P1 is present</td>
</tr>
<tr>
<td>KA</td>
<td>Affected if either P2 or P3 is present</td>
</tr>
<tr>
<td>NY</td>
<td>Affected if either P1, P2 or P3 is present</td>
</tr>
</tbody>
</table>

Each of the three groups from the simulated countries of AI, DA and KA collected 100 nuclear families diagnosed with the simulated disease while the NYC group collected 50 extended families with the disease. The summary of the disease model and the locations of the disease-related loci are given in Tables 3 & 4 respectively.
<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Major loci involved</th>
<th>Disease allele frequency</th>
<th>Inheritance at loci Epistatic models</th>
<th>Penetrance</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>D1</td>
<td>0.015</td>
<td>Dominant-Dominant</td>
<td>Penetrance of genotype is 0.6</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>D2</td>
<td>0.15</td>
<td>D2-D3</td>
<td>If D6 has allele 1, penetrance is 0.3, otherwise it is 0.6</td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>0.2</td>
<td>Recessive-Dominant</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>0.3</td>
<td>D3-D4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dominant-Recessive</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>D1</td>
<td>0.015</td>
<td>D1-D4</td>
<td>Penetrance of genotype is 1.0</td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>0.3</td>
<td>Dominant-Recessive</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>0.15</td>
<td>D2-D3</td>
<td>Penetrance of genotype is 0.4</td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>0.2</td>
<td>Dominant-Recessive</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. Location of disease related loci.

<table>
<thead>
<tr>
<th>Disease locus</th>
<th>Located between markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>C01R0052 and B01T0561</td>
</tr>
<tr>
<td>D2</td>
<td>B03T3067 and C04R0282</td>
</tr>
<tr>
<td>D3</td>
<td>B05T4136 and C05R0380</td>
</tr>
<tr>
<td>D4</td>
<td>C09R0765 and B09T8337</td>
</tr>
</tbody>
</table>
4.0 METHODS

4.1 OVERVIEW

The simulated dataset from GAW14 was analyzed in order to get a fair idea about how common methods of linkage analysis perform on complex disorders. Since gene interaction and heterogeneity had been incorporated into the genetic model of the simulated disorder in GAW14 and our main objective was to get a better understanding of gene interaction and heterogeneity in complex disorders, this dataset seemed to be very appropriate for the type of questions we wanted to address in complex disorders. As this set of simulated data mimicked a qualitative disorder, we could use methods which are employed in other complex qualitative disorders, such as nonparametric linkage scans, association studies of susceptibility regions, and conditional studies (conditioning on a previously identified susceptibility locus). We tried various methods of conditional analysis like stratification, redefinition, weighted analysis and logistic regression. In this simulated dataset, since it was possible to purchase additional flanking markers for regions of interest following an initial genome scan, we thought that it would be a good exercise to learn how to manage research funds efficiently.
4.2 NONPARAMETRIC LINKAGE ANALYSIS

We began with a standard nonparametric genome scan on the simulated data from replicate 1 with all the microsatellites and SNPs genome-screening markers – 1333 markers in all. We analyzed each of the four populations (Aipotu, Karangar, Danacaa, and NYC) independently. The software Merlin\(^1\) was used to perform the initial genome scans. This yielded linkage signals in four regions on different chromosomes (chromosomes 1, 3, 5, & 9) with suggestive to significant LOD scores with the highest LOD scores on chromosome 3 in all the four populations. In order to perform an in depth analysis on the regions of interest, we purchased sets of markers on chromosomes 1, 3, 5, and 9 flanking the regions with the highest LOD scores from the nonparametric analyses.
4.3 ANALYSES ON THE PURCHASED DATA

4.3.1 Linkage analysis and TDT Analysis (to find the associated alleles)

The additional markers purchased were incorporated into the existing data, and genome scans were repeated on the purchased data set. Minor changes were observed (mostly narrowing the width of the linkage peaks). Follow-up association-based results (transmission disequilibrium test – TDT) were generated with sib_tdt from the ASPEX package\textsuperscript{35} for single locus results; while multilocus TDT tests were done with TRANSMIT.\textsuperscript{7} The purpose of the follow-up association-based tests was to identify the over-transmitted alleles in the four regions with the highest linkage signals from the initial genome scan analysis. These association-based tests identified the over-transmitted alleles in the four regions with the highest linkage signals. During further analyses, the dataset for the NYC population could not be run properly as the analyses repeatedly crashed due to large family size. So, further analyses and comparisons were restricted to the other three populations.

4.3.2 Interaction Analyses

In order to investigate optimal methods for conditioning linkage analyses, we tried a number of methods. Conditional analysis in genetics can be used to describe the relationship between putative susceptibility genes. Complex disorders involve two or more genes acting together with or without environmental factors. Genome scans may be used to identify single
genes but may not be helpful in detecting the presence of additional genes involved in complex disorders that have an influence over the primary loci. Conditional analysis may be a handy tool in situations where there is a need to differentiate whether the effects are because of heterogeneity or because of interaction. Most methods in conditional analyses aim at stratifying families according to the linkage evidence at a specific locus and analyze them at a second region or genomewide.\textsuperscript{11, 47}

Initially it was Gurling et al\textsuperscript{29} who reported linkage by using a two-locus model. Smyth et al\textsuperscript{84} were one of the first to use a conditional method involving interaction between two loci in bipolar disorder based on the work done by Gurling et al\textsuperscript{29}. Kuida and Beier\textsuperscript{48} have shown that (in a murine model) polycystic kidney disease (PKD) progression in the juvenile cystic kidney (jck) mutation can be influenced by interacting modifiers and they localized one of these loci to chromosome 1. Using a chromosome 1 congenic strain, they improved the genetic analysis and mapped the interacting locus to proximal chromosome 4 with a highly significant lod score. Pierik et al\textsuperscript{71} studied gene-environment interactions in CD families and found evidence for an interaction between IBD4 and smoking. Linkage was only observed in CD families where at least one of the affected siblings was an active smoker at the time CD was diagnosed. For our analyses on the GAW14 simulated data, we tried some of the existing methods in conditional analyses and some new methods.

4.3.2.1 Stratification

Stratification has been a common procedure used in conditional analysis of complex disorders. The method helps in identifying new regions of linkage thereby supporting epistasis or heterogeneity. Many different ways of stratification have been tried so far in conditional analyses. Stratification by known genotypes, stratification by phenotypes, and stratification by
disease models are some examples. Shaw et al\textsuperscript{80} utilized this scheme to find interactive loci with the already determined locus 16q12 (NOD2/CARD15) in a genome-wide search for inflammatory bowel disease susceptibility loci. In this study, the group with the CARD15 variant had suggestive linkage results in 6p and 10p with LOD scores of 3.06 and 2.29 respectively. Hampe et al\textsuperscript{31} used stratification by known IBD genotypes. By stratifying on CARD15/NOD2 genotypes, they found evidence for a second IBD locus on chromosome 16p (the IBD8 locus).

We tried stratifying the families based on the overtransmitted alleles in the four above-mentioned regions looking for clues for interaction or heterogeneity. The stratification scheme used was designed to divide the initial data set into two groups – one with families of affected siblings homozygous for the allele of interest and the other, with families of affected siblings heterozygous for the same allele. Since some populations had affected siblings as either all homozygous or all heterozygous for the overtransmitted alleles (all affected individuals in the AI and KA populations were homozygous for allele 1 at B03T3056, and all affected individuals in the DA population were heterozygous for this allele in replicate 1), stratification analyses were not successful in yielding useful results in all populations. The stratification scheme was therefore modified to split each population into two groups – one (named ‘present’) with families of affected siblings who were homozygous for the allele of interest and the other (named ‘absent’) with the rest of the families in that population.

The software Merlin\textsuperscript{1} was used to carry out nonparametric linkage analysis in each of the stratified pedigree groups. The NYC families are all large extended families as opposed to the other centers, who all collected nuclear families. As Merlin was not able to handle these large pedigrees, Simwalk2 was used to perform nonparametric linkage analysis on the NYC population.
In order to comment on the significance of the results of the split in the LOD scores between the two groups in each population, we decided to do 1000 simulations using the software Allegro. For each simulation, a program was written to randomly split every population into two groups (without replacement), with the number of pedigrees in both the groups similar to the number of pedigrees in the stratified groups. The absolute differences in the LOD scores between these two stratified groups from 1000 simulations were taken into account to calculate the p-value for the LOD score difference obtained during the nonparametric linkage analysis for each stratified group.

4.3.2.2 Weighted analysis

Cox et al, in their study on non-insulin-dependent diabetes (NIDDM), describe an approach to assess the evidence for statistical interactions between unlinked regions that allows multipoint allele-sharing analysis to take the evidence for linkage at one region into account in assessing the evidence for linkage over the rest of the genome. Using this method, they show that the interaction of genes of chromosomes 2 (NIDDM1) and 15 (near CYP19) makes a contribution to susceptibility to type 2 diabetes in Mexican Americans from Starr County, Texas. They used two weighting schemes: in the weight 0-1 scheme, families with NPL scores of 0 or negative at NIDDM1 were assigned weight 0 and families with positive NPL scores were assigned weight 1; in the other scheme (weight 1-0), they assigned families weight 1 if their NPL score at NIDDM1 was 0 or negative and weight 0 if their NPL score at NIDDM1 was 0 or positive.

A conditional approach has been explored in asthma. Four regions were identified in a genomewide screen (the regions with the highest LOD scores). For each region, analyses were performed by conditioning on that region and searching the rest of the genome for evidence of
additional signals. Based on the evidence for linkage at these four regions, two weighting schemes were used to model a positive and a negative relationship between loci on different chromosomes. In the weight$_{0,1}$ model, a positive relationship (gene-gene interaction) was considered by assigning a weight of 0 to families with a LOD score of 0 or negative and a weight of 1 to families with positive LOD scores. In the weight$_{1,0}$ model (heterogeneity), a negative relationship was modeled. Here, families with negative linkage scores were given a weight of 1 and families with positive linkage scores or a LOD score of 0 were given a weight of 0.

McInnis et al$^{58}$ used this scheme to perform conditional analysis in bipolar disorder. Following a primary nonparametric genome-wide scan, they performed conditional analyses based on epistasis or heterogeneity, by weighting the families based on linkage results for the five regions with the highest linkage results. They reported seven potential interactions (four epistasis and three heterogeneity models) in which the NPL scores increased by at least 1.

We started with genome-wide linkage analyses on the dataset using Merlin conditioning on all the four loci with the highest nonparametric LOD scores. For each of the conditioned loci, weights of 1 or 0 were assigned to each family based on the LOD score at the conditional locus. A score of 1 was given to all the pedigrees which had a positive LOD score, and a score of 0 was given to all the pedigrees which had a negative LOD score or a LOD score of 0. The group with the LOD score of 1 was named ‘positive’ and the group with the LOD score of 0 was called ‘negative’ in our analyses. And thus we ended up in stratifying the pedigrees according to the presence or absence of linkage to specific regions. Non-parametric linkage analyses were again performed on the stratified pedigree subgroups using Merlin to look for additional positive signals on other regions of the genome.
Two conditional models were examined here – the epistatic model and the heterogeneity model. For the epistatic conditional model, nonparametric linkage analyses were performed on the subgroups with a positive LOD score. Any positive signal on some other region of the genome other than the conditioned locus was considered an interaction since this group already had positive signals for the conditioned locus. For the heterogeneity conditional model, nonparametric linkage analyses were performed on the other subgroups (which were assigned a score of 0). Since these groups did not have any linkage results initially, any positive signal in these groups was considered evidence of heterogeneity.

We also performed permutation tests for each conditional model in all three populations. In each simulation, a program was written to randomly choose a fixed number of pedigrees 1000 times, the number of pedigrees being similar to the number of pedigrees in the subgroups of the weighted analyses. The distribution of LOD scores from the simulations were taken into consideration to calculate the p-value for the LOD score obtained during the nonparametric analysis in each subgroup of the weighted analyses.

4.3.2.3 Redefinition

Redefinition is a concept that has been rarely tried so far by researchers in conditional analyses. Here, instead of splitting the families into groups, the phenotype (affection status) of the individuals is redefined based on the genotypes at regions with significant linkage signals in genomescans. Linkage analyses are repeated after redefining the phenotypes to see how the linkage signals change at the other loci. This redefinition scheme can be used to look for gene-gene interactions as well as gene-environment interactions.

In the redefinition scheme, we redefined the phenotype (affection status) of the individuals based on the genotypes at the four regions of interest mentioned above. The
The redefinition procedure was used in two different ways. At first, only individuals who were homozygous for an overtransmitted allele and were previously affected were classified as affected (the ‘homoz’ group). Next, only individuals who were heterozygous for the overtransmitted allele and were previously affected were classified as affected (the ‘heteroz’ group). Nonparametric linkage analyses were repeated after redefining the phenotypes to see how the linkage signals changed at the other loci.

4.3.2.4 Logistic Regression

Logistic regression falls under the class of statistical models called generalized linear models (GLM). Logistic regression can be used either to predict group membership or to understand the relationships and strengths among variables.

The model:

While logistic regression makes no assumption about the distribution of the independent or predictor variables, the dependent variable is dichotomous, that is, the dependent variable can take the value 1 with a probability of success $p$, or the value 0 with probability of failure $1-p$. This type of variable is called a binary variable.

For example, if $p$ = the probability of an outcome (e.g., occurrence of a disease) then a possible model could be of the form

$$E(p) = \beta_0 + \beta_1 X_1 + \ldots + \beta_{k-1} X_{k-1} = x \beta$$

where “$E$” stands for expected value, $\beta_0$ = intercept, $\beta_i$ to $\beta_{k-1}$ = coefficients.
However, the above model has the property that sometimes the right hand side (RHS) can be $<0$ or $>0$. (Recall that probabilities must be between 0 and 1, inclusive.) Since the values on the RHS in (1) can potentially take on values between $-\infty$ and $\infty$, both sides of the equation have to be transformed so that a very large (or small) value on the RHS of (1) maps into a value between 0 and 1, inclusive. One convenient transformation is that called the logistic or logit transformation of $p$. If we first note that the odds measure, $p/q = p/(1-p)$, has the property that $0 \leq p/(1-p) \leq \infty$, then we can see that if we take the natural log of the odds we have the property that $-\infty < \ln(p/(1-p)) < \infty$. Thus, instead of using the model represented in (1), we use

$$
\text{logit}(p) = \ln(p/(1-p)) = \ln(p/q) = \ln(\text{odds}) = \beta_0 + \beta_1 X_1 + \ldots + \beta_{k-1} X_{k-1}
$$

(2) 

North et al applied logistic regression to case-control association studies involving two causative loci, where the two susceptibility loci jointly influenced the risk of developing disease. In their work, simulated case-control samples that were generated assuming different two-locus models were used in the analyses using logistic regression.

We used the lrmodel program written by North to implement logistic regression modeling on our dataset. (The notations and formulas used by North et al have been used here to explain how the method works). Our initial nonparametric linkage analysis on the entire dataset yielded four susceptibility regions on different chromosomes – chr1, chr3, chr5, and chr9. When we considered all the possible two-locus interactions among these loci, we came up with six two-locus models for each of the four populations AI, DA, KA and NYC. In each case, a disease model is characterized based on the risk of having the disease conditional on the two-locus genotype. If $i$ and $j$ are the genotypes at the two loci (where $i, j = 0, 1$ or 2 corresponding to
genotypes 11, 12 and 22 at each locus), then the nine resulting (unphased) multilocus genotypes can be denoted by \( k \) (where \( k=3i+j \)). LRMODEL calculates the multilocus marker genotype frequencies for cases and controls by adding together the genotypings over the total number of cases and controls. If the penetrance for genotype \( k \) is termed as \( f_k \) and the frequency of genotype \( k \) in the population as \( g_k \), then the expected frequency of genotype \( k \) in a sample of affected cases is:

\[
P^p_k = g_k f_k / K
\]

and in a sample of unaffected controls is:

\[
P^u_k = g_k (1-f_k) / (1-K)
\]

where \( K \) is the overall disease prevalence.

A hypothetical dataset consisting of 1000 cases and 1000 controls is got by multiplying these expected frequencies by 1000. LRMODEL fits a sequence of analysis models by means of logistic regression. Each analysis model contains a subset of the full set of genetic effects. If “\( r \)” is the probability of a subset with the given genotype being a case rather a control, a full model for two loci can be given by:

\[
\log(r/1-r) = \mu + a_1 x_1 + d_1 z_1 + a_2 x_2 + d_2 z_2 + i_{aa} x_1 x_2 + i_{ad} x_1 z_2 + i_{da} z_1 x_2 + i_{dd} z_1 z_2
\] (3)

where \( x \) and \( z \) - dummy variables specific to each locus, defining additive and dominance effects

\( \mu \) - mean

\( a_1, a_2 \) - additive effects of the two loci

\( d_1, d_2 \) - dominance effects of the two loci

\( i_{aa}, i_{ad}, i_{da}, i_{dd} \) - interactive effects
The various combinations of these additive, dominance and interactive effects form the basis for the disease models used for the analysis as described by North et al. ADD1, ADD2 and ADD stand for models with additive effects, DOM1, DOM2 and DOM stand for models with dominance effects, and ADDINT, ADDDOM and DOMINT stand for models with interactive effects.

The method provides two measures of adequacy, the Likelihood Ratio Statistic (LRS) and Akaike’s Information Criterion (AIC).

\[
\text{LRS} = 2 \ln(L_2/L_1)
\]

\[
\text{AIC} = -2 \ln(L) + b \quad \{b \text{ is the number of free parameters in the model}\}
\]

The AIC measure is a substitute to the LRS for model assessment. A model with the lowest AIC value is regarded to have the best fit and most parsimonious.
5.0 RESULTS

5.1 INITIAL LINKAGE ANALYSIS

Nonparametric genome scans on the simulated data from replicate 1 yielded interesting NPL results for markers on chromosomes 1, 3, 5, and 9 with LOD scores greater than 1.5. A particular region on chromosome 3 showed significant linkage with suggestive (>2.0) or significant (>3.5) LOD scores in all populations. As we were considering these populations to be independent studies, we considered similar results in different populations reflective of replication. Based on these criteria, loci on chromosomes 1 and 3 are considered “confirmed”, while loci on chromosomes 5 and 9 are only “suggested” (Table 5).

Table 5. Results from genome scan with microsatellite and SNP genome scan markers.

<table>
<thead>
<tr>
<th>Population (# of individuals)</th>
<th>*LOD (cM)</th>
<th>Chr1</th>
<th>Chr3</th>
<th>Chr5</th>
<th>Chr9</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI (783)</td>
<td>1.72 (22)</td>
<td>3.85 (87.295)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DA (700)</td>
<td>6.48 (48.5)</td>
<td>2.15 (89.295)</td>
<td>1.97 (53.806)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KA (694)</td>
<td>5.89 (95.295)</td>
<td>5.19 (1.806)</td>
<td>5.18 (2.306)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NY (943)</td>
<td>2.4 (54)</td>
<td>2.73 (94.795)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Results are presented as npl LOD scores and locations at which the maximum score occurred. Only LOD scores > 1.5 are presented.
5.2 ANALYSES ON THE PURCHASED DATA

5.2.1 Linkage Analysis and Association Analysis

After purchasing relevant data sets based on the results of the primary analyses, genome scans were repeated on the purchased data set (after incorporating the additional markers into the existing data). The results are provided in Figure 6.

Figure 6. Comparison of LOD scores from npl analyses on purchased data – by chromosome and by population
Transmission disequilibrium tests on the purchased data set using ASPEX (sib_tdt) and TRANSMIT packages yielded significant single locus chi-squared scores for each population. The region on chromosome 3 (snp B03T3056) had the greatest chi-squared scores (ranging between 3.56 and 34.16) for all the populations. The TDT test also showed that in all these populations, for marker B03T3056 on chromosome 3, allele 1 was overtransmitted by a ratio of 2.5:1. Similar results were obtained when all four populations were combined. We also used TRANSMIT to test for multi-locus association between the genetic markers of interest on these chromosomes and the disease, to see if there were particular haplotypes which were associated. In so doing, we identified multi(2)-locus haplotypes which exhibited evidence of linkage to the disease phenotype, but with significance levels no greater than the single-locus tests. The overtransmitted alleles for the four regions with the highest linkage signals are listed in Table 6. Most of the conditioning in the future analyses was based on these overtransmitted alleles.

**Table 6.** Overtransmitted alleles for regions with highest linkage signals on the genome.

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Locus</th>
<th>Allele #</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B01T0558</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>B03T3056</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>B05T4140</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>B09T8333</td>
<td>2</td>
</tr>
</tbody>
</table>

The above mentioned analyses and tests were carried out on replicate 2, which yielded similar results.
5.2.2 Interaction Analysis

As one of our primary objectives was to comprehend how heterogeneity and interaction operate in complex disorders, further analyses were focused on conditional methods that could help detect heterogeneity and interaction in the complex disorder represented in GAW14 simulated data set. Also efforts were made to compare how these methods perform with respect to one another. As described earlier, four different conditional methods were employed and nonparametric linkage analyses were carried out to see how the linkage signals behaved at the conditioned loci in different populations under different methods. Also, in order to examine the effect of sample size, the conditional analyses were repeated on bigger samples by combining some of the replicates. Replicates 1 and 2 were combined to double the sample size and replicates 1 through 5 were combined to produce an even bigger sample size.

5.2.2.1 Stratification

The pedigrees in each population were split into two groups based on any affected sibling being homozygous for an overtransmitted allele as described earlier. The results of the nonparametric analyses from each of the stratified groups by population are laid out in Tables 7, 8, 9 and 10. These tables also contain results from initial nonparametric linkage analysis on the full sets of pedigrees for a quick comparison.
Table 7. Nonparametric analyses in the stratified groups – Replicate 1.

<table>
<thead>
<tr>
<th>Based on</th>
<th>Pop</th>
<th># of aff. Fam.</th>
<th>chr 1 LOD scores</th>
<th>chr 3 LOD scores</th>
<th>chr 5 LOD scores</th>
<th>chr 9 LOD scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>46.841-</td>
<td>89.725-</td>
<td>1.757-</td>
<td>1.721-</td>
</tr>
<tr>
<td>chr 1</td>
<td>ai</td>
<td>present</td>
<td>38</td>
<td>0.02</td>
<td>0.71</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>62</td>
<td>0.13</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>0.69</td>
<td>4.24</td>
<td>0.92</td>
</tr>
<tr>
<td>da</td>
<td>present</td>
<td>46</td>
<td>4.35</td>
<td>1.72</td>
<td>-0.02</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>absent</td>
<td>54</td>
<td>1.69</td>
<td>2.56</td>
<td>0.64</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>100</td>
<td>5.56</td>
<td>3.02</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td>ka</td>
<td>present</td>
<td>35</td>
<td>0.02</td>
<td>1.61</td>
<td>0.4</td>
<td>3.4</td>
</tr>
<tr>
<td></td>
<td>absent</td>
<td>65</td>
<td>0.47</td>
<td>4.18</td>
<td>5.5</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>100</td>
<td>5.39</td>
<td>5.8</td>
<td>5.53</td>
<td>4.1</td>
</tr>
<tr>
<td>chr 3</td>
<td>ai</td>
<td>present</td>
<td>71</td>
<td>0.63</td>
<td>7.64</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>29</td>
<td>no result</td>
<td>no result</td>
<td>no result</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>100</td>
<td>0.69</td>
<td>4.24</td>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td>ka</td>
<td>present</td>
<td>69</td>
<td>0</td>
<td>9.01</td>
<td>2.53</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>absent</td>
<td>31</td>
<td>no result</td>
<td>no result</td>
<td>no result</td>
<td>no result</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>100</td>
<td>0.39</td>
<td>5.8</td>
<td>5.53</td>
<td>4.1</td>
</tr>
<tr>
<td>chr 5</td>
<td>ai</td>
<td>present</td>
<td>69</td>
<td>1.4</td>
<td>2.69</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>31</td>
<td>0.02</td>
<td>2.33</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>100</td>
<td>0.69</td>
<td>4.24</td>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td>da</td>
<td>present</td>
<td>27</td>
<td>3.71</td>
<td>0.56</td>
<td>-0.05</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>absent</td>
<td>73</td>
<td>4.06</td>
<td>1.97</td>
<td>0.61</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>100</td>
<td>5.56</td>
<td>3.02</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td>ka</td>
<td>present</td>
<td>84</td>
<td>0.22</td>
<td>4.66</td>
<td>5.92</td>
<td>4.92</td>
</tr>
<tr>
<td></td>
<td>absent</td>
<td>16</td>
<td>0.21</td>
<td>0.85</td>
<td>0.22</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>100</td>
<td>0.39</td>
<td>5.8</td>
<td>5.53</td>
<td>4.1</td>
</tr>
<tr>
<td>chr 9</td>
<td>ai</td>
<td>present</td>
<td>26</td>
<td>0.31</td>
<td>0.01</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>74</td>
<td>0.18</td>
<td>6.58</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>100</td>
<td>0.69</td>
<td>4.24</td>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td>da</td>
<td>present</td>
<td>33</td>
<td>1.2</td>
<td>2</td>
<td>0.23</td>
<td>-0.72</td>
</tr>
<tr>
<td></td>
<td>absent</td>
<td>67</td>
<td>6.1</td>
<td>0.88</td>
<td>0.09</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>100</td>
<td>5.56</td>
<td>3.02</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td>ka</td>
<td>present</td>
<td>48</td>
<td>1.7</td>
<td>1.35</td>
<td>3.5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>absent</td>
<td>52</td>
<td>0</td>
<td>5.5</td>
<td>2.35</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>100</td>
<td>0.39</td>
<td>5.8</td>
<td>5.53</td>
<td>4.1</td>
</tr>
</tbody>
</table>
Table 8. Nonparametric analyses in the stratified groups – Replicate 2.

<table>
<thead>
<tr>
<th>Based on</th>
<th>Pop</th>
<th># of aff.</th>
<th>LOD scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Fam.</td>
<td>chr 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>46.841-</td>
</tr>
<tr>
<td>chr1</td>
<td>ai</td>
<td>present</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td>chr3</td>
<td>ai</td>
<td>present</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td>chr5</td>
<td>ai</td>
<td>present</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td>chr9</td>
<td>ai</td>
<td>present</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 9. Nonparametric analyses in the stratified groups – Combined replicates 1 and 2.

<table>
<thead>
<tr>
<th>Based on</th>
<th>Pop.</th>
<th># of aff. Fam.</th>
<th>LOD scores</th>
<th>chr 1</th>
<th>chr 3</th>
<th>chr 5</th>
<th>chr 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46.841-48.451 cM</td>
<td>89.725-94.881 cM</td>
<td>1.757-2.181 cM</td>
<td>1.721-2.154 cM</td>
</tr>
<tr>
<td>chr1</td>
<td>ai</td>
<td>present</td>
<td>-0.01</td>
<td>1.69</td>
<td>0.07</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>2.32</td>
<td>2.5</td>
<td>3.33</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>1.3</td>
<td>4.19</td>
<td>2.49</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>5.76</td>
<td>2.34</td>
<td>0.17</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>3.41</td>
<td>3.03</td>
<td>0.3</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>8.82</td>
<td>5.35</td>
<td>0.47</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>-0.13</td>
<td>2.27</td>
<td>0.95</td>
<td>5.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>0.97</td>
<td>3.9</td>
<td>10.29</td>
<td>3.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>0.37</td>
<td>6.16</td>
<td>10.42</td>
<td>8.76</td>
<td></td>
</tr>
<tr>
<td>chr3</td>
<td>ai</td>
<td>present</td>
<td>0.46</td>
<td>2.26</td>
<td>2.56</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>1.03</td>
<td>2.01</td>
<td>0.28</td>
<td>1.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>1.3</td>
<td>4.19</td>
<td>2.49</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>5.49</td>
<td>3.51</td>
<td>0.6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>3.37</td>
<td>1.84</td>
<td>0</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>8.82</td>
<td>5.35</td>
<td>0.47</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>-0.13</td>
<td>4.76</td>
<td>7.37</td>
<td>2.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>2.85</td>
<td>1.52</td>
<td>3.05</td>
<td>8.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>0.37</td>
<td>6.16</td>
<td>10.42</td>
<td>8.76</td>
<td></td>
</tr>
<tr>
<td>chr5</td>
<td>ai</td>
<td>present</td>
<td>0.85</td>
<td>2.07</td>
<td>0.06</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>0.49</td>
<td>2.13</td>
<td>3.95</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>1.3</td>
<td>4.19</td>
<td>2.49</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>3.76</td>
<td>2.93</td>
<td>-1.03</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>5.1</td>
<td>2.47</td>
<td>3.9</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>8.82</td>
<td>5.35</td>
<td>0.47</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>0.04</td>
<td>4.53</td>
<td>2.7</td>
<td>4.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>0.4</td>
<td>2.18</td>
<td>7.91</td>
<td>4.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>0.37</td>
<td>6.16</td>
<td>10.42</td>
<td>8.76</td>
<td></td>
</tr>
<tr>
<td>chr9</td>
<td>ai</td>
<td>present</td>
<td>2.19</td>
<td>5.57</td>
<td>2.01</td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>0.13</td>
<td>0.57</td>
<td>0.77</td>
<td>2.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>1.3</td>
<td>4.19</td>
<td>2.49</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>7.38</td>
<td>1.44</td>
<td>0.13</td>
<td>-1.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>2.34</td>
<td>4.12</td>
<td>0.36</td>
<td>2.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>8.82</td>
<td>5.35</td>
<td>0.47</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>0</td>
<td>5.06</td>
<td>3.79</td>
<td>-0.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>0.53</td>
<td>2.25</td>
<td>6.64</td>
<td>14.84</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>0.37</td>
<td>6.16</td>
<td>10.42</td>
<td>8.76</td>
<td></td>
</tr>
<tr>
<td>Based on</td>
<td>Pop</td>
<td># of aff. Fam.</td>
<td>chr 1 46.841-48.451 cM</td>
<td>chr 3 89.725-94.881 cM</td>
<td>chr 5 1.757-2.181 cM</td>
<td>chr 9 1.721-2.154 cM</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>----------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>chr1</td>
<td>ai</td>
<td>present</td>
<td>193 1.41</td>
<td>5.51</td>
<td>0.95</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>307 6.36</td>
<td>9.08</td>
<td>6.03</td>
<td>3.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 7.47</td>
<td>14.59</td>
<td>6.42</td>
<td>4.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>209 13.02</td>
<td>6.4</td>
<td>0.27</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>291 20.44</td>
<td>9.88</td>
<td>0.88</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 33.45</td>
<td>16.28</td>
<td>1.1</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>185 -0.24</td>
<td>3.63</td>
<td>2.81</td>
<td>9.71</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>315 1.82</td>
<td>6.25</td>
<td>22.97</td>
<td>6.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 0.66</td>
<td>9.88</td>
<td>23.21</td>
<td>15.82</td>
<td></td>
</tr>
<tr>
<td>chr3</td>
<td>ai</td>
<td>present</td>
<td>319 2.85</td>
<td>10.7</td>
<td>6.57</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>181 5.39</td>
<td>4.01</td>
<td>0.73</td>
<td>5.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 7.47</td>
<td>14.59</td>
<td>6.42</td>
<td>4.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>324 22.15</td>
<td>11.46</td>
<td>0.74</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>176 11.3</td>
<td>5.01</td>
<td>0.35</td>
<td>1.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 33.45</td>
<td>16.28</td>
<td>1.1</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>325 -0.01</td>
<td>8.6</td>
<td>20.23</td>
<td>2.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>175 2.36</td>
<td>1.81</td>
<td>4.2</td>
<td>21.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 0.66</td>
<td>9.88</td>
<td>23.21</td>
<td>15.82</td>
<td></td>
</tr>
<tr>
<td>chr5</td>
<td>ai</td>
<td>present</td>
<td>229 6.68</td>
<td>7.76</td>
<td>0.06</td>
<td>1.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>271 1.75</td>
<td>6.85</td>
<td>10.18</td>
<td>2.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 7.47</td>
<td>14.59</td>
<td>6.42</td>
<td>4.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>226 13.6</td>
<td>8.57</td>
<td>-1.93</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>274 19.97</td>
<td>7.83</td>
<td>6.54</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 33.45</td>
<td>16.28</td>
<td>1.1</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>209 0.13</td>
<td>6.3</td>
<td>6.4</td>
<td>9.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>291 0.57</td>
<td>4.13</td>
<td>17.27</td>
<td>6.81</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 0.66</td>
<td>9.88</td>
<td>23.21</td>
<td>15.82</td>
<td></td>
</tr>
<tr>
<td>chr9</td>
<td>ai</td>
<td>present</td>
<td>184 4.82</td>
<td>10.85</td>
<td>3.72</td>
<td>-0.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>316 3.08</td>
<td>5.43</td>
<td>2.9</td>
<td>8.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 7.47</td>
<td>14.59</td>
<td>6.42</td>
<td>4.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>present</td>
<td>224 20.2</td>
<td>7.9</td>
<td>0.04</td>
<td>-1.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>276 13.97</td>
<td>8.41</td>
<td>1.5</td>
<td>5.76</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 33.45</td>
<td>16.28</td>
<td>1.1</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>present</td>
<td>162 0.21</td>
<td>8.05</td>
<td>14.11</td>
<td>-0.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>absent</td>
<td>338 0.45</td>
<td>3.43</td>
<td>10.46</td>
<td>26.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500 0.66</td>
<td>9.88</td>
<td>23.21</td>
<td>15.82</td>
<td></td>
</tr>
</tbody>
</table>
In replicate 1, while stratifying based on allele 1 for the locus B03T3056 on chromosome 3, we found that there were no informative affected pairs for the disease in the Aipotu and Karangar populations in the ‘absent’ groups. As a result, there was no output for these groups in these two populations. For the same replicate, in the Danacaa population, all the affected individuals were heterozygous for allele 1 (locus B03T3056). So, it was not possible to have any families in the ‘present’ group to carry out the stratification. It was noted that the LOD scores in most of the stratified groups were unevenly distributed in all the populations.

Tables 11, 12, 13 and 14 contain the results of the simulations done on the stratified groups. LOD scores greater than 5.0 with significant p-values were seen in a few stratified groups.

**Table 11. Simulations on stratified groups – Replicate 1.**

<table>
<thead>
<tr>
<th>Based on</th>
<th>chr 1</th>
<th>chr 3</th>
<th>chr 5</th>
<th>chr 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>diff</td>
<td>p-value</td>
<td>diff</td>
<td>p-value</td>
</tr>
<tr>
<td>chr 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>0.11</td>
<td>0.964</td>
<td>0.71</td>
<td>0.732</td>
</tr>
<tr>
<td>da</td>
<td>2.66</td>
<td>0.128</td>
<td>0.84</td>
<td>0.633</td>
</tr>
<tr>
<td>ka</td>
<td>0.45</td>
<td>0.807</td>
<td>2.57</td>
<td>0.093</td>
</tr>
<tr>
<td>chr 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>1.38</td>
<td>0.704</td>
<td>0.36</td>
<td>0.917</td>
</tr>
<tr>
<td>da</td>
<td>0.35</td>
<td>0.941</td>
<td>0.41</td>
<td>0.926</td>
</tr>
<tr>
<td>ka</td>
<td>0.01</td>
<td>0.984</td>
<td>3.81</td>
<td>0.013</td>
</tr>
<tr>
<td>chr 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>0.13</td>
<td>0.956</td>
<td>6.57</td>
<td>0</td>
</tr>
<tr>
<td>da</td>
<td>4.9</td>
<td>0.024</td>
<td>1.12</td>
<td>0.672</td>
</tr>
<tr>
<td>ka</td>
<td>1.7</td>
<td>0.307</td>
<td>4.15</td>
<td>0.016</td>
</tr>
</tbody>
</table>

P-values indicative of heterogeneity and interaction are printed in bold and underlined respectively. NR- No results; ND- Not done
### Table 12. Simulations on stratified groups – Replicate 2.

<table>
<thead>
<tr>
<th>Based on</th>
<th>chr 1</th>
<th>chr 3</th>
<th>chr 5</th>
<th>chr 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>diff</td>
<td>p-value</td>
<td>diff</td>
<td>p-value</td>
</tr>
<tr>
<td>chr 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>1.72</td>
<td>0.069</td>
<td>0.71</td>
<td>0.355</td>
</tr>
<tr>
<td>da</td>
<td>1.16</td>
<td>0.563</td>
<td>0.31</td>
<td>0.9</td>
</tr>
<tr>
<td>ka</td>
<td>0.03</td>
<td>0.981</td>
<td>0.32</td>
<td>0.813</td>
</tr>
<tr>
<td>chr 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>0.31</td>
<td>0.712</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>da</td>
<td>0.27</td>
<td>0.817</td>
<td>1.16</td>
<td>0.327</td>
</tr>
<tr>
<td>ka</td>
<td>2.22</td>
<td>0.031</td>
<td>0.66</td>
<td>0.358</td>
</tr>
<tr>
<td>chr 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>0.32</td>
<td>0.664</td>
<td>0.06</td>
<td>0.946</td>
</tr>
<tr>
<td>da</td>
<td>3.65</td>
<td>0.007</td>
<td>0.95</td>
<td>0.56</td>
</tr>
<tr>
<td>ka</td>
<td>0.19</td>
<td>0.876</td>
<td>1.01</td>
<td>0.381</td>
</tr>
<tr>
<td>chr 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>0.25</td>
<td>0.782</td>
<td>0.63</td>
<td>0.473</td>
</tr>
<tr>
<td>da</td>
<td>2.12</td>
<td>0.503</td>
<td>1.67</td>
<td>0.637</td>
</tr>
<tr>
<td>ka</td>
<td>0.66</td>
<td>0.834</td>
<td>2.02</td>
<td>0.37</td>
</tr>
</tbody>
</table>

P-values indicative of heterogeneity and interaction are printed in bold and underlined respectively. NR- No results; ND- Not done

### Table 13. Simulations on stratified groups – Combined replicates 1 and 2.

<table>
<thead>
<tr>
<th>Based on</th>
<th>chr 1</th>
<th>chr 3</th>
<th>chr 5</th>
<th>chr 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>diff</td>
<td>p-value</td>
<td>diff</td>
<td>p-value</td>
</tr>
<tr>
<td>chr 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>2.33</td>
<td>0.235</td>
<td>0.81</td>
<td>0.726</td>
</tr>
<tr>
<td>da</td>
<td>2.35</td>
<td>0.187</td>
<td>0.69</td>
<td>0.737</td>
</tr>
<tr>
<td>ka</td>
<td>1.1</td>
<td>0.595</td>
<td>1.77</td>
<td>0.372</td>
</tr>
<tr>
<td>chr 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>0.57</td>
<td>0.908</td>
<td>0.25</td>
<td>0.957</td>
</tr>
<tr>
<td>da</td>
<td>2.12</td>
<td>0.503</td>
<td>1.67</td>
<td>0.637</td>
</tr>
<tr>
<td>ka</td>
<td>2.98</td>
<td>0.361</td>
<td>3.24</td>
<td>0.302</td>
</tr>
<tr>
<td>chr 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>0.36</td>
<td>0.914</td>
<td>0.06</td>
<td>0.986</td>
</tr>
<tr>
<td>da</td>
<td>1.34</td>
<td>0.476</td>
<td>0.46</td>
<td>0.824</td>
</tr>
<tr>
<td>ka</td>
<td>0.36</td>
<td>0.897</td>
<td>2.35</td>
<td>0.267</td>
</tr>
<tr>
<td>chr 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>2.06</td>
<td>0.29</td>
<td>5</td>
<td>0.002</td>
</tr>
<tr>
<td>da</td>
<td>5.04</td>
<td>0.007</td>
<td>2.68</td>
<td>0.136</td>
</tr>
<tr>
<td>ka</td>
<td>0.53</td>
<td>0.768</td>
<td>2.81</td>
<td>0.107</td>
</tr>
</tbody>
</table>

P-values indicative of heterogeneity and interaction are printed in bold and underlined respectively. NR- No results; ND- Not done
Table 14. Simulations on stratified groups – Combined replicates 1 through 5.

<table>
<thead>
<tr>
<th>Based on</th>
<th>chr 1</th>
<th>chr 3</th>
<th>chr 5</th>
<th>chr 9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>diff</td>
<td>p-value</td>
<td>diff</td>
<td>p-value</td>
</tr>
<tr>
<td>chr 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>4.95</td>
<td>0.315</td>
<td>3.57</td>
<td>0.523</td>
</tr>
<tr>
<td>da</td>
<td>7.42</td>
<td>0.806</td>
<td>3.48</td>
<td>0.924</td>
</tr>
<tr>
<td>ka</td>
<td>2.06</td>
<td>0.327</td>
<td>2.62</td>
<td>0.212</td>
</tr>
<tr>
<td>chr 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>2.54</td>
<td>0.362</td>
<td>6.69</td>
<td>0.028</td>
</tr>
<tr>
<td>da</td>
<td>10.85</td>
<td>0.073</td>
<td>6.45</td>
<td>0.295</td>
</tr>
<tr>
<td>ka</td>
<td>2.37</td>
<td>0.472</td>
<td>6.79</td>
<td>0.002</td>
</tr>
<tr>
<td>chr 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>4.93</td>
<td>0.165</td>
<td>0.91</td>
<td>0.847</td>
</tr>
<tr>
<td>da</td>
<td>6.37</td>
<td>0.749</td>
<td>0.74</td>
<td>0.975</td>
</tr>
<tr>
<td>ka</td>
<td>0.44</td>
<td>0.82</td>
<td>2.17</td>
<td>0.298</td>
</tr>
<tr>
<td>chr 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai</td>
<td>1.74</td>
<td>0.838</td>
<td>5.42</td>
<td>0.276</td>
</tr>
<tr>
<td>da</td>
<td>6.23</td>
<td>0.766</td>
<td>0.51</td>
<td>0.979</td>
</tr>
<tr>
<td>ka</td>
<td>0.24</td>
<td>0.93</td>
<td>4.62</td>
<td>0.046</td>
</tr>
</tbody>
</table>

P-values indicative of heterogeneity and interaction are printed in bold and underlined respectively. NR- No results; ND- Not done

For replicate 1, simulations were not possible for the locus on chromosome 3 (B03T3056) as either there were inadequate informative affected pairs for the disease in the AI and KA populations, or there were no homozygous affected individuals for the same locus in the DA population. While comparing the results of the simulations with the expected number of interactions in the GAW dataset, we observed that in all the replicates the number of false positives was greater than the number of true positives.

5.2.2.2 Weighted Analysis

After splitting the families into those which had positive LOD scores and those which had negative LOD scores or a LOD score of 0 at the conditioned loci from an initial genomscan, follow-up analyses were carried out on these subgroups as described below. A total of 24
additional nonparametric linkage analyses were done on these subgroups for each set of replicates (single or combined). The results of these linkage analyses are presented in tables 15, 16, 17 and 18.

To determine the significance of the results of the conditional analyses, we performed 1000 simulations for each of the subgroups. The results of the simulations are presented in tables 15, 16, 17 and 18. With the help of the simulation analyses, we found out that this method also did not work effectively on the given simulated dataset in detecting interaction or heterogeneity (very low percentage of true positives).
Table 15. Nonparametric analyses and simulations in the weighted groups – Replicate 1.

<table>
<thead>
<tr>
<th>Based on</th>
<th>Pop</th>
<th>Sub-group</th>
<th># of aff.</th>
<th>LOD scores</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fam.</td>
<td></td>
<td>chr 1</td>
<td>chr 3</td>
<td>chr 5</td>
<td>chr 9</td>
<td></td>
</tr>
<tr>
<td>chr1</td>
<td>ai</td>
<td>positive</td>
<td>52</td>
<td>0</td>
<td>3.58</td>
<td>0.181</td>
<td>0.29</td>
<td>0.9</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47</td>
<td>1</td>
<td>1.14</td>
<td>0.748</td>
<td>0.58</td>
<td>0.972</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
<td>4.24</td>
<td>0.92</td>
<td></td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr3</td>
<td>ai</td>
<td>positive</td>
<td>59</td>
<td>0</td>
<td>4.81</td>
<td>0.14</td>
<td>5.08</td>
<td>0.009</td>
<td>2.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40</td>
<td>1</td>
<td>-0.95</td>
<td>1.25</td>
<td>0.892</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>1</td>
<td>5.8</td>
<td>5.53</td>
<td></td>
<td></td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr5</td>
<td>ai</td>
<td>positive</td>
<td>53</td>
<td>0.19</td>
<td>1.97</td>
<td>0.773</td>
<td>10.89</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47</td>
<td>0.8</td>
<td>2.5</td>
<td>0.217</td>
<td>-1.23</td>
<td>1</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>0.69</td>
<td>4.24</td>
<td>0.92</td>
<td></td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr9</td>
<td>ai</td>
<td>positive</td>
<td>42</td>
<td>0.43</td>
<td>1.82</td>
<td>0.712</td>
<td>0.1</td>
<td>1</td>
<td>8.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>58</td>
<td>0.41</td>
<td>2.59</td>
<td>0.283</td>
<td>0.96</td>
<td>0.935</td>
<td>-1.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td>0.69</td>
<td>4.24</td>
<td>0.92</td>
<td></td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P-values indicative of heterogeneity and interaction are printed in bold and underlined respectively. NR- No results; ND- Not done
<table>
<thead>
<tr>
<th>Based on</th>
<th>Pop</th>
<th>Subgroup</th>
<th># of aff. Fam.</th>
<th>LOD scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>chr 1 46.841-48.451 cM</td>
<td>chr 3 89.725-94.881 cM</td>
</tr>
<tr>
<td>chr1</td>
<td>ai</td>
<td>positive</td>
<td>47</td>
<td>10.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>53</td>
<td>-2.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>0.73</td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>57</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>43</td>
<td>-2.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>3.58</td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>47</td>
<td>8.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>53</td>
<td>-4.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>-0.01</td>
</tr>
<tr>
<td>chr3</td>
<td>ai</td>
<td>positive</td>
<td>45</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>55</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>0.73</td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>60</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>40</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>3.58</td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>51</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>49</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>-0.01</td>
</tr>
<tr>
<td>chr5</td>
<td>ai</td>
<td>positive</td>
<td>57</td>
<td>1.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>43</td>
<td>-0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>0.73</td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>42</td>
<td>2.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>58</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>3.58</td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>63</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>37</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>-0.01</td>
</tr>
<tr>
<td>chr9</td>
<td>ai</td>
<td>positive</td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>100</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>0.73</td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>100</td>
<td>3.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>3.58</td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>0</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>100</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>100</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

P-values indicative of heterogeneity and interaction are printed in bold and underlined respectively. NR- No results; ND- Not done
<table>
<thead>
<tr>
<th>Based on</th>
<th>Pop</th>
<th>Sub-group</th>
<th># of aff. Fam.</th>
<th>LOD scores</th>
<th>LOD scores</th>
<th>LOD scores</th>
<th>LOD scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>chr 1</td>
<td>chr 3</td>
<td>chr 5</td>
<td>chr 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46.841-48.451 cM</td>
<td>89.725-94.881 cM</td>
<td>1.757-2.181 cM</td>
<td>1.721-2.154 cM</td>
</tr>
<tr>
<td>chr1</td>
<td>ai</td>
<td>positive</td>
<td>97</td>
<td>19.87</td>
<td>3.91</td>
<td>0.205</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>103</td>
<td>-4.48</td>
<td>0.95</td>
<td>0.917</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>1.3</td>
<td>4.19</td>
<td>2.49</td>
<td>0.744</td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>121</td>
<td>28.03</td>
<td>2.88</td>
<td>0.999</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>79</td>
<td>-4.01</td>
<td>2.5</td>
<td>0.972</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>8.82</td>
<td>5.35</td>
<td>0.47</td>
<td>1.12</td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>97</td>
<td>19.25</td>
<td>2.02</td>
<td>0.874</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>103</td>
<td>-5.4</td>
<td>1</td>
<td>0.95</td>
<td>0.974</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>0.37</td>
<td>6.16</td>
<td>10.42</td>
<td>8.76</td>
</tr>
<tr>
<td>chr3</td>
<td>ai</td>
<td>positive</td>
<td>104</td>
<td>3.09</td>
<td>26.33</td>
<td>0</td>
<td>2.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>96</td>
<td>-0.01</td>
<td>-2.28</td>
<td>1</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>1.3</td>
<td>4.19</td>
<td>2.49</td>
<td>0.874</td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>114</td>
<td>4.69</td>
<td>26.87</td>
<td>0</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>86</td>
<td>4.18</td>
<td>-3.76</td>
<td>1</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>8.82</td>
<td>5.35</td>
<td>0.47</td>
<td>1.12</td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>110</td>
<td>0.3</td>
<td>26.4</td>
<td>0</td>
<td>5.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>90</td>
<td>0.09</td>
<td>-3.6</td>
<td>1</td>
<td>4.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>0.37</td>
<td>6.16</td>
<td>10.42</td>
<td>8.76</td>
</tr>
<tr>
<td>chr5</td>
<td>ai</td>
<td>positive</td>
<td>113</td>
<td>1.3</td>
<td>1.99</td>
<td>0.908</td>
<td>23.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>87</td>
<td>0.16</td>
<td>2.23</td>
<td>0.128</td>
<td>-2.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>1.3</td>
<td>4.19</td>
<td>2.49</td>
<td>1.46</td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>92</td>
<td>4.28</td>
<td>1.66</td>
<td>0.997</td>
<td>19.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>108</td>
<td>4.56</td>
<td>3.8</td>
<td>0.866</td>
<td>-5.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>8.82</td>
<td>5.35</td>
<td>0.47</td>
<td>1.12</td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>124</td>
<td>0.37</td>
<td>3.17</td>
<td>0.388</td>
<td>26.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>76</td>
<td>0.04</td>
<td>3.25</td>
<td>0.031</td>
<td>-2.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>0.37</td>
<td>6.16</td>
<td>10.42</td>
<td>8.76</td>
</tr>
<tr>
<td>chr9</td>
<td>ai</td>
<td>positive</td>
<td>0</td>
<td>NR</td>
<td>ND</td>
<td>ND</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>200</td>
<td>1.3</td>
<td>ND</td>
<td>ND</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>1.3</td>
<td>4.19</td>
<td>2.49</td>
<td>1.12</td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>0</td>
<td>NR</td>
<td>ND</td>
<td>ND</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>200</td>
<td>8.82</td>
<td>ND</td>
<td>ND</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>8.82</td>
<td>5.35</td>
<td>0.47</td>
<td>0.1</td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>0</td>
<td>NR</td>
<td>ND</td>
<td>ND</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>200</td>
<td>0.37</td>
<td>ND</td>
<td>ND</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>200</td>
<td>0.37</td>
<td>6.16</td>
<td>10.42</td>
<td>8.76</td>
</tr>
</tbody>
</table>

P-values indicative of heterogeneity and interaction are printed in bold and underlined respectively. NR- No results; ND- Not done.
Table 18. Nonparametric analyses in the weighted groups – Combined replicates 1 through 5.

<table>
<thead>
<tr>
<th>Based on</th>
<th>Pop</th>
<th>Subgroup</th>
<th># of aff. Fam.</th>
<th>chr 1 46.841-48.451 cM</th>
<th>sim. p-value</th>
<th>chr 3 89.725-94.881 cM</th>
<th>sim. p-value</th>
<th>chr 5 1.757-2.181 cM</th>
<th>sim. p-value</th>
<th>chr 9 1.721-2.154 cM</th>
<th>sim. p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1</td>
<td>ai</td>
<td>positive</td>
<td>258</td>
<td>57.6</td>
<td>0</td>
<td>7.68</td>
<td>0.104</td>
<td>4.28</td>
<td>0.915</td>
<td>2.05</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>242</td>
<td>-9.96</td>
<td>1</td>
<td>6.93</td>
<td>0.579</td>
<td>2.27</td>
<td>1</td>
<td>2.46</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>7.47</td>
<td>14.59</td>
<td>6.42</td>
<td>4.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>332</td>
<td>75.93</td>
<td>0</td>
<td>10.42</td>
<td>0.999</td>
<td>0.37</td>
<td>1</td>
<td>0.04</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>168</td>
<td>-7.37</td>
<td>1</td>
<td>5.86</td>
<td>1</td>
<td>0.85</td>
<td>1</td>
<td>2.29</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>33.45</td>
<td>16.28</td>
<td>1.1</td>
<td>1.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>226</td>
<td>47.83</td>
<td>0</td>
<td>3.13</td>
<td>0.478</td>
<td>7.04</td>
<td>0.005</td>
<td>5.99</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>274</td>
<td>-13.34</td>
<td>1</td>
<td>7.11</td>
<td>0</td>
<td>16.64</td>
<td>0</td>
<td>9.99</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>0.66</td>
<td>9.88</td>
<td>23.21</td>
<td>15.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr3</td>
<td>ai</td>
<td>positive</td>
<td>272</td>
<td>5.45</td>
<td>0.661</td>
<td>66.93</td>
<td>0</td>
<td>5.91</td>
<td>0.513</td>
<td>1.06</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>228</td>
<td>2.25</td>
<td>1</td>
<td>-4.94</td>
<td>1</td>
<td>1.25</td>
<td>1</td>
<td>3.89</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>7.47</td>
<td>14.59</td>
<td>6.42</td>
<td>4.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>283</td>
<td>18.86</td>
<td>0.091</td>
<td>68.07</td>
<td>0</td>
<td>0.53</td>
<td>1</td>
<td>0.18</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>217</td>
<td>14.64</td>
<td>0.861</td>
<td>-8.44</td>
<td>1</td>
<td>1.57</td>
<td>1</td>
<td>1.2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>33.45</td>
<td>16.28</td>
<td>1.1</td>
<td>1.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>252</td>
<td>0.46</td>
<td>0.999</td>
<td>60.43</td>
<td>0</td>
<td>13.12</td>
<td>0</td>
<td>5.29</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>248</td>
<td>0.22</td>
<td>1</td>
<td>-10.07</td>
<td>1</td>
<td>10.17</td>
<td>0</td>
<td>11.03</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>0.66</td>
<td>9.88</td>
<td>23.21</td>
<td>15.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr5</td>
<td>ai</td>
<td>positive</td>
<td>279</td>
<td>4.48</td>
<td>0.939</td>
<td>8.8</td>
<td>0.064</td>
<td>59.18</td>
<td>0</td>
<td>0.64</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>221</td>
<td>2.99</td>
<td>0.997</td>
<td>5.84</td>
<td>0.764</td>
<td>-5.97</td>
<td>1</td>
<td>5.2</td>
<td>0.858</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>7.47</td>
<td>14.59</td>
<td>6.42</td>
<td>4.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>231</td>
<td>13.63</td>
<td>0.226</td>
<td>6.24</td>
<td>1</td>
<td>50.57</td>
<td>0</td>
<td>1.79</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>269</td>
<td>19.86</td>
<td>0.697</td>
<td>10.26</td>
<td>1</td>
<td>-13.43</td>
<td>1</td>
<td>0.03</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>33.45</td>
<td>16.28</td>
<td>1.1</td>
<td>1.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>303</td>
<td>0.45</td>
<td>1</td>
<td>6.63</td>
<td>0.006</td>
<td>67.38</td>
<td>0</td>
<td>5.72</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>197</td>
<td>0.21</td>
<td>1</td>
<td>3.26</td>
<td>0.116</td>
<td>-6.61</td>
<td>1</td>
<td>11.19</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>0.66</td>
<td>9.88</td>
<td>23.21</td>
<td>15.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr9</td>
<td>ai</td>
<td>positive</td>
<td>0</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>500</td>
<td>7.47</td>
<td>ND</td>
<td>14.59</td>
<td>6.42</td>
<td>ND</td>
<td>4.48</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>7.47</td>
<td>14.59</td>
<td>6.42</td>
<td>4.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>da</td>
<td></td>
<td>positive</td>
<td>0</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>500</td>
<td>33.45</td>
<td>ND</td>
<td>16.28</td>
<td>1.1</td>
<td>ND</td>
<td>1.13</td>
<td>ND</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>33.45</td>
<td>16.28</td>
<td>1.1</td>
<td>1.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ka</td>
<td></td>
<td>positive</td>
<td>0</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>ND</td>
<td>NR</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative</td>
<td>500</td>
<td>0.66</td>
<td>ND</td>
<td>9.88</td>
<td>23.21</td>
<td>ND</td>
<td>15.82</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>0.66</td>
<td>9.88</td>
<td>23.21</td>
<td>15.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P-values indicative of heterogeneity and interaction are printed in bold and underlined respectively. NR- No results; ND- Not done
5.2.2.3 Redefinition

The results of the analyses based on redefining the phenotypes of the original dataset are summarized in tables 19, 20, 21 and 22. A quick glance at the nature of the distribution of the alleles at the conditioned loci from this table showed that except for the marker B03T3056 on chromosome 3, every other marker that we conditioned on had their alleles distributed in a heterozygous fashion in the majority of the cases in all the populations. Only marker B03T3056 on chromosome 3 had some tendency of being distributed in the homozygous fashion in most of the cases in all the populations.
<table>
<thead>
<tr>
<th>Based on</th>
<th>Pop</th>
<th>Sub-group</th>
<th># of affecteds</th>
<th>chr 1 LOD scores</th>
<th>chr 3 LOD scores</th>
<th>chr 5 LOD scores</th>
<th>chr 9 LOD scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46.841-48.451 cM</td>
<td>89.725-94.881 cM</td>
<td>1.757-2.181 cM</td>
<td>1.721-2.154 cM</td>
</tr>
<tr>
<td>chr1</td>
<td>ai</td>
<td>homoz</td>
<td>78</td>
<td>0.45</td>
<td>0.72</td>
<td>0.31</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>145</td>
<td>2.2</td>
<td>1.92</td>
<td>3.43</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>281</td>
<td>0.69</td>
<td>4.24</td>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td>da</td>
<td>homoz</td>
<td></td>
<td>97</td>
<td><strong>7.46</strong></td>
<td>0</td>
<td>-0.4</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>118</td>
<td>3.04</td>
<td>2.19</td>
<td>0.48</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>258</td>
<td>5.56</td>
<td>3.02</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td>ka</td>
<td>homoz</td>
<td></td>
<td>56</td>
<td>1.22</td>
<td>0.48</td>
<td>0.21</td>
<td>3.52</td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>135</td>
<td>3</td>
<td>4.23</td>
<td>4.21</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>242</td>
<td>0.39</td>
<td>5.8</td>
<td>5.53</td>
<td>4.1</td>
</tr>
<tr>
<td>chr3</td>
<td>ai</td>
<td>homoz</td>
<td>160</td>
<td>0.1</td>
<td><strong>7.64</strong></td>
<td>0.65</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>111</td>
<td>1.73</td>
<td>3.41</td>
<td>0</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>281</td>
<td>0.69</td>
<td>4.24</td>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td>da</td>
<td>homoz</td>
<td></td>
<td>0</td>
<td>no result</td>
<td>no result</td>
<td>no result</td>
<td>no result</td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>107</td>
<td>2.42</td>
<td>1.69</td>
<td>0.15</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>258</td>
<td>5.56</td>
<td>3.02</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td>ka</td>
<td>homoz</td>
<td></td>
<td>136</td>
<td>0</td>
<td><strong>9.01</strong></td>
<td>2.53</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>87</td>
<td>1.12</td>
<td>1.3</td>
<td>2.01</td>
<td>3.77</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>242</td>
<td>0.39</td>
<td>5.8</td>
<td>5.53</td>
<td>4.1</td>
</tr>
<tr>
<td>chr5</td>
<td>ai</td>
<td>homoz</td>
<td>67</td>
<td>0.37</td>
<td>0.89</td>
<td>4.3</td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>130</td>
<td>0.01</td>
<td>0.63</td>
<td>3.01</td>
<td>1.48</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>281</td>
<td>0.69</td>
<td>4.24</td>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td>da</td>
<td>homoz</td>
<td></td>
<td>50</td>
<td>0.67</td>
<td>0.33</td>
<td>2.58</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>116</td>
<td>0.77</td>
<td>0.29</td>
<td>1.09</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>258</td>
<td>5.56</td>
<td>3.02</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td>ka</td>
<td>homoz</td>
<td></td>
<td>50</td>
<td>1.87</td>
<td>1.84</td>
<td>2.39</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>102</td>
<td>-0.06</td>
<td>0.43</td>
<td>4.35</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>242</td>
<td>0.39</td>
<td>5.8</td>
<td>5.53</td>
<td>4.1</td>
</tr>
<tr>
<td>chr9</td>
<td>ai</td>
<td>homoz</td>
<td>55</td>
<td>0.28</td>
<td>0.06</td>
<td>0.01</td>
<td>1.51</td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>142</td>
<td>-0.07</td>
<td>2.79</td>
<td>0</td>
<td>1.36</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>281</td>
<td>0.69</td>
<td>4.24</td>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td>da</td>
<td>homoz</td>
<td></td>
<td>59</td>
<td>0.02</td>
<td>0.35</td>
<td>1.84</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>108</td>
<td>0.38</td>
<td>1.61</td>
<td>0.37</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>258</td>
<td>5.56</td>
<td>3.02</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td>ka</td>
<td>homoz</td>
<td></td>
<td>94</td>
<td>0.26</td>
<td>0.13</td>
<td>1.98</td>
<td><strong>7.31</strong></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td></td>
<td>104</td>
<td>0</td>
<td>3.22</td>
<td>1.6</td>
<td>3.18</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td></td>
<td>242</td>
<td>0.39</td>
<td>5.8</td>
<td>5.53</td>
<td>4.1</td>
</tr>
</tbody>
</table>

LOD scores indicative of heterogeneity are printed in bold.
Table 20. Nonparametric analyses in the redefined groups – Replicate 2.

<table>
<thead>
<tr>
<th>Based on</th>
<th>Pop</th>
<th>Sub-group</th>
<th># of affecteds</th>
<th>LOD scores chr1</th>
<th>LOD scores chr3</th>
<th>LOD scores chr5</th>
<th>LOD scores chr9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>46.841-8.451 cM</td>
<td>89.725-94.881 cM</td>
<td>1.757-2.181 cM</td>
<td>1.721-2.154 cM</td>
</tr>
<tr>
<td>chr1 ai</td>
<td>homoz</td>
<td>77</td>
<td>1.01</td>
<td>0.78</td>
<td>1.44</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>119</td>
<td>2.77</td>
<td>0.07</td>
<td>0.76</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>256</td>
<td>0.73</td>
<td>0.57</td>
<td>1.85</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da homoz</td>
<td>81</td>
<td>5.1</td>
<td>0.9</td>
<td>0.19</td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>116</td>
<td>1.3</td>
<td>0.28</td>
<td>0.35</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>253</td>
<td>3.58</td>
<td>3.01</td>
<td>0.13</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka homoz</td>
<td>68</td>
<td>1.52</td>
<td>0.15</td>
<td>0.52</td>
<td>1.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>132</td>
<td>0.73</td>
<td>0.12</td>
<td>2.49</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>258</td>
<td>-0.01</td>
<td>1.32</td>
<td>5.52</td>
<td>4.92</td>
<td></td>
</tr>
<tr>
<td>chr3 ai</td>
<td>homoz</td>
<td>125</td>
<td>-0.05</td>
<td>1.99</td>
<td>0.78</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>93</td>
<td>0.85</td>
<td>2.2</td>
<td>0.65</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>256</td>
<td>0.73</td>
<td>0.57</td>
<td>1.85</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da homoz</td>
<td>127</td>
<td>0.1</td>
<td>5.37</td>
<td>0.19</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>119</td>
<td>2.63</td>
<td>2.22</td>
<td>-0.13</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>253</td>
<td>3.58</td>
<td>3.01</td>
<td>0.13</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka homoz</td>
<td>145</td>
<td>-0.33</td>
<td>4.31</td>
<td>5.78</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>79</td>
<td>0.44</td>
<td>0.53</td>
<td>1</td>
<td>1.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>258</td>
<td>-0.01</td>
<td>1.32</td>
<td>5.52</td>
<td>4.92</td>
<td></td>
</tr>
<tr>
<td>chr5 ai</td>
<td>homoz</td>
<td>47</td>
<td>0.02</td>
<td>0.01</td>
<td>0.62</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>117</td>
<td>0.19</td>
<td>0.15</td>
<td>3.22</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>256</td>
<td>0.73</td>
<td>0.57</td>
<td>1.85</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da homoz</td>
<td>53</td>
<td>0.61</td>
<td>1.85</td>
<td>2.43</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>130</td>
<td>2.63</td>
<td>0.74</td>
<td>2.13</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>253</td>
<td>3.58</td>
<td>3.01</td>
<td>0.13</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka homoz</td>
<td>36</td>
<td>-0.01</td>
<td>0.07</td>
<td>2.23</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>140</td>
<td>0.1</td>
<td>0.56</td>
<td>4.02</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>258</td>
<td>-0.01</td>
<td>1.32</td>
<td>5.52</td>
<td>4.92</td>
<td></td>
</tr>
<tr>
<td>chr9 ai</td>
<td>homoz</td>
<td>71</td>
<td>0.15</td>
<td>-0.03</td>
<td>0.27</td>
<td>5.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>128</td>
<td>1.1</td>
<td>0.1</td>
<td>0.7</td>
<td>1.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>256</td>
<td>0.73</td>
<td>0.57</td>
<td>1.85</td>
<td>1.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>da homoz</td>
<td>63</td>
<td>0.57</td>
<td>0.58</td>
<td>0.07</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>115</td>
<td>0.21</td>
<td>0.38</td>
<td>0</td>
<td>1.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>253</td>
<td>3.58</td>
<td>3.01</td>
<td>0.13</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ka homoz</td>
<td>83</td>
<td>0.37</td>
<td>-0.06</td>
<td>1.75</td>
<td>6.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>heteroz</td>
<td>116</td>
<td>-0.01</td>
<td>0.49</td>
<td>0.58</td>
<td>2.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>258</td>
<td>-0.01</td>
<td>1.32</td>
<td>5.52</td>
<td>4.92</td>
<td></td>
</tr>
</tbody>
</table>

LOD scores indicative of heterogeneity and interaction are printed in bold and underlined respectively.
### Table 21. Nonparametric analyses in the redefined groups – Combined replicates 1 and 2.

<table>
<thead>
<tr>
<th>Based on</th>
<th>Pop</th>
<th>Sub-group</th>
<th># of affecteds</th>
<th>LOD scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>chr 1 46.841-48.451 cM</td>
<td>chr 3 89.725-94.881 cM</td>
</tr>
<tr>
<td>chr1</td>
<td>ai</td>
<td>homoz</td>
<td>155</td>
<td>1.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>264</td>
<td>4.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>537</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>homoz</td>
<td>179</td>
<td><strong>12.25</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>233</td>
<td>3.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>511</td>
<td>8.82</td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>homoz</td>
<td>124</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>267</td>
<td>3.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>0.37</td>
</tr>
<tr>
<td>chr3</td>
<td>ai</td>
<td>homoz</td>
<td>285</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>204</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>537</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>homoz</td>
<td>271</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>228</td>
<td>4.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>511</td>
<td>8.82</td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>homoz</td>
<td>281</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>166</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>0.37</td>
</tr>
<tr>
<td>chr5</td>
<td>ai</td>
<td>homoz</td>
<td>114</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>247</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>537</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>homoz</td>
<td>103</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>246</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>511</td>
<td>8.82</td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>homoz</td>
<td>85</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>242</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>0.37</td>
</tr>
<tr>
<td>chr9</td>
<td>ai</td>
<td>homoz</td>
<td>126</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>270</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>537</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>da</td>
<td>homoz</td>
<td>122</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>223</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>511</td>
<td>8.82</td>
</tr>
<tr>
<td></td>
<td>ka</td>
<td>homoz</td>
<td>175</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>heteroz</td>
<td>220</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
<td>500</td>
<td>0.37</td>
</tr>
</tbody>
</table>

LOD scores indicative of heterogeneity and interaction are printed in bold and underlined respectively.
Table 22. Nonparametric analyses in the redefined groups – Combined replicates 1 through 5.

<table>
<thead>
<tr>
<th>Based of</th>
<th>Pop Subgroup</th>
<th># of affecteds</th>
<th>LOD scores (chromosome/cM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>chr 1 46.841-48.451 cM</td>
</tr>
<tr>
<td>chr1</td>
<td>ai homoz</td>
<td>365</td>
<td>11.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>654</td>
<td>11.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1332</td>
<td>7.47</td>
</tr>
<tr>
<td></td>
<td>da homoz</td>
<td>423</td>
<td>23.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>592</td>
<td>17.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1258</td>
<td>33.45</td>
</tr>
<tr>
<td></td>
<td>ka homoz</td>
<td>326</td>
<td>5.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>658</td>
<td>5.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1272</td>
<td>0.66</td>
</tr>
<tr>
<td>chr3</td>
<td>ai homoz</td>
<td>711</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>504</td>
<td>4.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1332</td>
<td>7.47</td>
</tr>
<tr>
<td></td>
<td>da homoz</td>
<td>641</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>574</td>
<td>13.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1258</td>
<td>33.45</td>
</tr>
<tr>
<td></td>
<td>ka homoz</td>
<td>659</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>454</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1272</td>
<td>0.66</td>
</tr>
<tr>
<td>chr5</td>
<td>ai homoz</td>
<td>270</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>618</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1332</td>
<td>7.47</td>
</tr>
<tr>
<td></td>
<td>da homoz</td>
<td>239</td>
<td>2.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>631</td>
<td>13.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1258</td>
<td>33.45</td>
</tr>
<tr>
<td></td>
<td>ka homoz</td>
<td>224</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>628</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1272</td>
<td>0.66</td>
</tr>
<tr>
<td>chr9</td>
<td>ai homoz</td>
<td>319</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>654</td>
<td>1.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1332</td>
<td>7.47</td>
</tr>
<tr>
<td></td>
<td>da homoz</td>
<td>263</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>596</td>
<td>9.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1258</td>
<td>33.45</td>
</tr>
<tr>
<td></td>
<td>ka homoz</td>
<td>379</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>599</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1272</td>
<td>0.66</td>
</tr>
</tbody>
</table>

LOD scores indicative of heterogeneity and interaction are printed in bold and underlined respectively.
5.2.2.4 Logistic Regression

As one of the methods of conditioning, a logistic regression approach was implemented as described previously. For each of the four populations, logistic regression analysis was performed on six two-locus interaction models. The summary of the results is displayed in table 23.

The table gives the AIC values for the dominant disease model and also the AIC difference. The AIC difference is calculated by taking the difference between the least AIC score for that interaction and the AIC score for a particular model. The model with the least AIC value is considered to have the best fit.

As most of the interaction models incorporated into the simulated data set included a dominance effect at least at one of the two interacting loci, we concentrated on the ‘DOM’ model in interpreting the output from the logistic regression analysis using the program LRMODEL. The other models in the output either talk about effects at single loci or about models not present in our data set.
Table 23. Logistic Regression analyses using the program LRModel.

<table>
<thead>
<tr>
<th>Pop</th>
<th>1&amp;3</th>
<th>1&amp;5</th>
<th>1&amp;9</th>
<th>3&amp;5</th>
<th>3&amp;9</th>
<th>5&amp;9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AI</td>
<td>AIC</td>
<td>AIC Diff</td>
<td>AIC</td>
<td>AIC Diff</td>
<td>AIC</td>
</tr>
<tr>
<td>Replicate 1 (DOM model)</td>
<td>AI</td>
<td>210.28</td>
<td>4.18</td>
<td>227.48</td>
<td>5.05</td>
<td>225.33</td>
</tr>
<tr>
<td></td>
<td>DA</td>
<td>165.70</td>
<td>1.19</td>
<td>175.04</td>
<td>3.96</td>
<td>176.62</td>
</tr>
<tr>
<td></td>
<td>KA</td>
<td>138.09</td>
<td>0</td>
<td>148.13</td>
<td>3.83</td>
<td>146.74</td>
</tr>
<tr>
<td></td>
<td>NY</td>
<td>51.01</td>
<td>1.75</td>
<td>85.20</td>
<td>5.93</td>
<td>83.77</td>
</tr>
<tr>
<td>Replicate 2 (DOM model)</td>
<td>AI</td>
<td>177.55</td>
<td>2.29</td>
<td>193.44</td>
<td>5.95</td>
<td>194.26</td>
</tr>
<tr>
<td></td>
<td>DA</td>
<td>183.27</td>
<td>2.95</td>
<td>194.54</td>
<td>7.05</td>
<td>189.24</td>
</tr>
<tr>
<td></td>
<td>KA</td>
<td>168.88</td>
<td>3.44</td>
<td>180.98</td>
<td>6.46</td>
<td>181.31</td>
</tr>
<tr>
<td></td>
<td>NY</td>
<td>85.6</td>
<td>3.69</td>
<td>88.72</td>
<td>5.72</td>
<td>86.52</td>
</tr>
<tr>
<td>Combined replicates 1&amp;2 (DOM model)</td>
<td>AI</td>
<td>380.26</td>
<td>0.95</td>
<td>413.59</td>
<td>3.16</td>
<td>412.68</td>
</tr>
<tr>
<td></td>
<td>DA</td>
<td>341.22</td>
<td>1.07</td>
<td>361.03</td>
<td>4.24</td>
<td>359.64</td>
</tr>
<tr>
<td></td>
<td>KA</td>
<td>301.74</td>
<td>0</td>
<td>324.78</td>
<td>3.61</td>
<td>323.62</td>
</tr>
<tr>
<td>Combined replicates 1 thru 5 (DOM model)</td>
<td>AI</td>
<td>974.06</td>
<td>3.71</td>
<td>1031.76</td>
<td>4.64</td>
<td>1035.6</td>
</tr>
<tr>
<td></td>
<td>DA</td>
<td>844.17</td>
<td>0</td>
<td>876.95</td>
<td>3.96</td>
<td>876.93</td>
</tr>
<tr>
<td></td>
<td>KA</td>
<td>867.4</td>
<td>3.04</td>
<td>899.89</td>
<td>5.41</td>
<td>896.8</td>
</tr>
<tr>
<td></td>
<td>NY</td>
<td>358.55</td>
<td>4.38</td>
<td>414.52</td>
<td>7.55</td>
<td>410.0</td>
</tr>
<tr>
<td></td>
<td>NY</td>
<td>95.84</td>
<td>5.18</td>
<td>158.92</td>
<td>6.22</td>
<td>156.59</td>
</tr>
</tbody>
</table>

For each dataset the analysis models which are within an AIC difference of 2 compared with the best model (AIC difference =0) are highlighted.
6.0 DISCUSSION

The stratification of pedigrees based on the homozygosity of alleles at the four loci B01T0558, B03T3056, B05T4140 and B09T8333 reflected the unequal distribution of these genotypes among different populations. The results of the analyses on the stratified pedigree groups indicated the possibility of false positive results in many populations, which led us to carry out simulation tests. As suggested, following simulation tests, the stratification method detected a very low percentage of true positives at a type I error (\(\alpha\)) level of 0.01. When we look at how the method performed for each of the four populations separately, the results are not significant. For the DA population, the method performed overly conservative because it did not detect any interaction at all (with neither true positives nor false positives). For both the AI and the KA populations, false positive rates were greater than true positive rates.

In the weighted analysis, there was a more even distribution of the genotypes between the ‘positive’ and ‘negative’ groups as opposed to the uneven representation in the stratified analyses. Nonparametric linkage analyses on the subgroups yielded very significant signals in all the populations at the four loci on chromosomes 1, 3, 5 and 9. But careful comparison of those interactions with the GAW ‘answers’ showed that most of those were false positive signals.

Both the redefinition method and the logistic regression method have a population-based approach instead of a family-based approach (as in stratification method or weighted analysis). In the redefinition method, individuals were considered as cases only if they had a particular
combination of alleles at the loci they were being conditioned on. In the logistic regression method, irrespective of the genotypes at the two interested loci, cases were those individuals who were originally affected in the populations. Logistic regression analysis was performed for each of the six two-locus interaction models possible in each of the four populations. Also logistic regression was the only analysis which was able to handle the huge pedigree sizes of the New York City population. The results from these analyses showed a very low power in the detection of the expected interactions. Comparing with each other, the logistic regression method seemed to be more meaningful than the redefinition method.

During the process of this study, we learned that every single method used had its advantages and disadvantages. Some analyses were easy to perform with a very short running time but the results were hard to interpret. The results of logistic regression are a good example of this problem (Table 23). While others took an eternity to run, the results were very clear (refer to tables 11 through 18 for the results of simulations on AI, DA and KA populations). Also, we understood that all the weaknesses of a method should be addressed and taken care of in order to obtain better results. This was true when we discovered that increasing the sample size alone did not increase the power in our analyses (Tables 7 through 22), while the other inadequacies of the methods were not addressed.

Researchers have always struggled to find the appropriate cut-off for LOD scores in interpreting their results even with the availability of guidelines on the significance of LOD scores by Lander and Kruglyak\textsuperscript{51}. In this light, we find the weighted analysis method more convincing since it groups all the families with any positive LOD score into one set for the conditional analysis, thereby minimizing the chances of missing any relevant information from the results. A quick glance at Table 24 will tell us that the weighted analysis method did better in
detecting more true positives in three of the four methods used. Also, the subgroups have a more even distribution of pedigrees (or individuals) in the weighted analysis as compared to other methods, which we attribute to the low cut-off LOD score of the weighted analysis method.

By and large, as evident from the results from the methods used, it is possible to say that these methods were successful in detecting heterogeneity more than interactions. In some cases, the signals for heterogeneity may be interpreted as suggestive of allelic heterogeneity rather than locus heterogeneity. For example, while conditioning on chromosome 9 (B09T8333) in Replicate 2 in the stratification method (Table 12), we get significant p-values in DA and KA populations at the same locus, which can be suggestive of allelic heterogeneity.
7.0 CONCLUSION

The main idea of this project was to study various methods of analyses in a dataset simulating a complex disease and comment on how these methods perform in detecting interaction or heterogeneity. A genome scan was successfully carried out on the simulated dataset of GAW14. Transmission Disequilibrium Test was used to identify loci with overtransmitted alleles. Further analyses were performed conditioning on these overtransmitted alleles using stratification method, redefinition method, weighted analysis method and logistic regression. The methods that we used did not prove to be very effective in detecting interaction or heterogeneity. Each of the methods we tried provided more false positive than true positive results (Table 24).

We believe that a number of reasons contributed to this fact as mentioned earlier in the background section. In our simulated dataset, incomplete penetrance, naïve assumption of the marker for conditioning, very narrow range of flanking markers and genetic heterogeneity (possibly allelic heterogeneity) are considered important factors in not achieving a high degree of performance.
Table 24. Comparison of results of conditional analyses with the GAW ‘answers’.

<table>
<thead>
<tr>
<th>Stratification</th>
<th>Weighted Analysis</th>
<th>Redefinition</th>
<th>Logistic Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exp.</td>
<td>Observed</td>
<td>Exp.</td>
</tr>
<tr>
<td></td>
<td>TP</td>
<td>FP</td>
<td>TP</td>
</tr>
<tr>
<td></td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>Replicate 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>(8.3)</td>
<td>(25)</td>
<td>(25)</td>
</tr>
<tr>
<td>Replicate 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Combined replicates 1 and 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Combined replicates 1 through 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

We also calculated the prevalence of interactions in the GAW14 simulated dataset from the disease allele frequencies (Table 25). This gave us a pretty clear picture of the low prevalence of interactions in this simulated dataset, which explained the inability of the methods to detect interactions. To improve the effectiveness of these methods we suggest that future investigators address these specific issues. Considering haplotype studies instead of targeting single markers, and increasing the range of the flanking markers around regions of high LOD scores could be some of the solutions to the problems encountered so far in this simulated dataset.
### Table 25. Prevalence of interactions.

<table>
<thead>
<tr>
<th>Population</th>
<th>Interaction</th>
<th>Model</th>
<th>Prevalence of interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AI</strong></td>
<td>Chromosomes 1</td>
<td>3</td>
<td>D/D</td>
</tr>
<tr>
<td></td>
<td>Disease allele 1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allele frequency 0.015</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chromosomes 1</strong></td>
<td>9</td>
<td>D/R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disease allele 1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allele frequency 0.015</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td><strong>Chromosomes 3</strong></td>
<td>5</td>
<td>D/R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disease allele 1</td>
<td>2</td>
<td>or</td>
</tr>
<tr>
<td></td>
<td>Allele frequency 0.15</td>
<td>0.2</td>
<td>or</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>or</td>
</tr>
<tr>
<td><strong>Chromosomes 5</strong></td>
<td>9</td>
<td>D/R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disease allele 2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allele frequency 0.2</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td><strong>DA</strong></td>
<td>Chromosomes 1</td>
<td>3</td>
<td>D/D</td>
</tr>
<tr>
<td></td>
<td>Disease allele 1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allele frequency 0.015</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KA</strong></td>
<td>Chromosomes 1</td>
<td>9</td>
<td>D/R</td>
</tr>
<tr>
<td></td>
<td>Disease allele 1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allele frequency 0.015</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td><strong>Chromosomes 3</strong></td>
<td>5</td>
<td>D/R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disease allele 1</td>
<td>2</td>
<td>or</td>
</tr>
<tr>
<td></td>
<td>Allele frequency 0.15</td>
<td>0.2</td>
<td>or</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>or</td>
</tr>
<tr>
<td><strong>Chromosomes 5</strong></td>
<td>9</td>
<td>D/R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disease allele 2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allele frequency 0.2</td>
<td>0.3</td>
<td></td>
</tr>
</tbody>
</table>
8.0 FUTURE STUDIES

The main objective of our study was to focus on the comparison of various methods of linkage analyses and conditional analyses in a simulated dataset of a complex disease. Even though interaction and heterogeneity were incorporated as main components into the simulated dataset of GAW14, our methods were not effective in detecting the same efficiently. By using more efficient strategies to carefully choose the regions on the genome for conditional analyses, future studies on this dataset could provide better results using the same methods used in this study.
APPENDIX A

SHELL SCRIPT FOR NON-PARAMETRIC LINKAGE ANALYSIS

#!/bin/tcsh -f

# C-shell file name: script.sh

#----------------------------------------------
#       Input file names:
#        Locus file:            datain.dat
#        Pedigree files:         ai.dat da.dat ka.dat ny.dat
#        Map file:              map.dat
#----------------------------------------------

mkdir Strat Redef Weight Logist

cd Strat

# STRATIFICATION ANALYSES

# Starting with AI population

# Stratifying based on chr1 locus B01T0558 - allele 1 (columns 41 & 42)

mkdir AI DA KA NY

mkdir AI_results DA_results KA_results NY_results

cd AI

mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1
mkdir Present Absent

cd Present

    cp ../../../ai.dat .
    cp ../../../datain.dat .
    cp ../../../map.dat .

    awk '$3 > 0' ai.dat > tmp1.dat
    awk '$6 > 1' tmp1.dat > tmp2.dat
    awk '$41 < 2' tmp2.dat > tmp3.dat
    awk '$42 < 2' tmp3.dat > tmp4.dat
    awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat
    rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file ai.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)  
#  awk "$1 == "$ped" ai.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl ai.dat pedlist1.dat pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_pre_01.lod

awk '/B01T0558/' merlin_out.01 > ai_pre_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_pre_03.lod

awk '/B03T3056/' merlin_out.03 > ai_pre_03.lod
# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_pre_05.lod

awk '/B05T4140/' merlin_out.05 > ai_pre_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_pre_09.lod

awk '/B09T8333/' merlin_out.09 > ai_pre_09.lod

cd ../../Absent

cp ../../../ai.dat .
cp ../../../datain.dat .
cp ../../../map.dat .

# compare ai.dat with pedin.dat in the directory-present, then retain those
# lines that are not in pedin.dat and write it to a file called pedin.dat

cp ./Present/pedin.dat pedin1.dat

cat ./pedin1.dat ./ai.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

 mega2 <<< end_of_input

2
dat
0
27
y
4
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_abs_01.lod

awk '/B01T0558/' merlin_out.01 > ai_abs_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_abs_03.lod

awk '/B03T3056/' merlin_out.03 > ai_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_abs_05.lod

awk '/B05T4140/' merlin_out.05 > ai_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_abs_09.lod
awk '/B09T8333/' merlin_out.09 > ai_abs_09.lod

cd ../..

# Move into the next : Chr3/

cd Chr3

# Repeating all the steps done for Chr1

mkdir Present Absent

cd Present

cp ../../../../ai.dat .
cp ../../../../datain.dat .
cp ../../../../map.dat .
awk '$3 > 0' ai.dat > tmp1.dat
awk '$6 > 1' tmp1.dat > tmp2.dat
awk '$117 < 2' tmp2.dat > tmp3.dat
awk '$118 < 2' tmp3.dat > tmp4.dat
awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat
rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file ai.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:

# (1) code in this file to loop over pedigrees in pedlist.dat:
# foreach ped (`cat pedlist.dat`)
#   awk '$1 == '$ped'' ai.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl ai.dat pedlist1.dat pedin.dat

# running mega2

    mega2 << end_of_input

2

dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
awk '/B01T0558/' merlin_out.01 > ai_pre_01.lod

Awk '/B03T3056/' merlin_out.03 > ai_pre_03.lod

Awk '/B05T4140/' merlin_out.05 > ai_pre_05.lod

Awk '/B09T8333/' merlin_out.09 > ai_pre_09.lod

cd ../Absent

cp ../../ai.dat .
cp ../../datain.dat .
cp ../../map.dat .
cp ../Present/pedin.dat pedin1.dat
cat ./pedin1.dat ./ai.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

71
dat

27

y

4

1

8

0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_abs_01.lod

    awk '/B01T0558/' merlin_out.01 > ai_abs_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_abs_03.lod

    awk '/B03T3056/' merlin_out.03 > ai_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_abs_05.lod

    awk '/B05T4140/' merlin_out.05 > ai_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_abs_09.lod

    awk '/B09T8333/' merlin_out.09 > ai_abs_09.lod

cd ../..

# Move into the next: Chr5

cd Chr5

# Repeat steps

mkdir Present Absent

cd Present

    cp ../../../ai.dat .
    cp ../../../datain.dat .
    cp ../../../map.dat .
    awk '$3 > 0' ai.dat > tmp1.dat
    awk '$6 > 1' tmp1.dat > tmp2.dat
    awk '$165 < 2' tmp2.dat > tmp3.dat
    awk '$166 < 2' tmp3.dat > tmp4.dat
    awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat
    rm tmp*
# if the number in the pedlist.dat matches the number in the first field
# of the file ai.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
#  (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)  
#   awk "$1 == "$ped" ai.dat >> pedin.dat  
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl ai.dat pedlist1.dat pedin.dat

# running mega2

    mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_pre_01.lod

awk '/B01T0558/' merlin_out.01 > ai_pre_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_pre_03.lod

awk '/B03T3056/' merlin_out.03 > ai_pre_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_pre_05.lod

awk '/B05T4140/' merlin_out.05 > ai_pre_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_pre_09.lod

awk '/B09T8333/' merlin_out.09 > ai_pre_09.lod

cd ../Absent
cp ../../../ai.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
cp ../../../Present/pedin.dat pedin1.dat
cat ./pedin1.dat ./ai.dat | sort | uniq -u > pedin.dat
    awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_abs_01.lod
awk '/B01T0558/' merlin_out.01 > ai_abs_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_abs_03.lod

awk '/B03T3056/' merlin_out.03 > ai_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_abs_05.lod

awk '/B05T4140/' merlin_out.05 > ai_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_abs_09.lod

awk '/B09T8333/' merlin_out.09 > ai_abs_09.lod

cd ../..

# Move into the next : Chr9

cd Chr9

# Repeat steps

mkdir Present Absent

cd Present

cp ../../../ai.dat .
cp ../../../../datain.dat .
cp ../../../../map.dat .
awk '$3 > 0' ai.dat > tmp1.dat
awk '$6 > 1' tmp1.dat > tmp2.dat
awk '$191 < 2' tmp2.dat > tmp3.dat
awk '$192 < 2' tmp3.dat > tmp4.dat
awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat
rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file ai.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)
#  awk '$1 == '$ped'' ai.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

pull_pedigrees.pl ai.dat pedlist1.dat pedin.dat

# running mega2
mega2 << end_of_input
2
dat
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_pre_01.lod

awk '/B01T0558/' merlin_out.01 > ai_pre_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_pre_03.lod

awk '/B03T3056/' merlin_out.03 > ai_pre_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_pre_05.lod
awk '/B05T4140/' merlin_out.05 > ai_pre_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_pre_09.lod

awk '/B09T8333/' merlin_out.09 > ai_pre_09.lod

cd ../Absent

cp ../../ai.dat .

cp ../../datain.dat .

cp ../../map.dat .

cp ../Present/pedin.dat pedin1.dat

cat ./pedin1.dat ./ai.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 <<< end_of_input

2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38

end_of_input
#!/merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_abs_01.lod

awk '/B01T0558/' merlin_out.01 > ai_abs_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_abs_03.lod

awk '/B03T3056/' merlin_out.03 > ai_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_abs_05.lod

awk '/B05T4140/' merlin_out.05 > ai_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_abs_09.lod

awk '/B09T8333/' merlin_out.09 > ai_abs_09.lod

# Move up by three levels

cd ../../../
# Move into the AI_results directory and copy the results only

cd AI_results
mkdir Chr1 Chr3 Chr5 Chr9
cd Chr1
cp ../../../AI/Chr1/Present/pedlist1.dat .
cp ../../../AI/Chr1/Present/ai_pre* .
cp ../../../AI/Chr1/Absent/pedlist2.dat .
cp ../../../AI/Chr1/Absent/ai_abs* .

cd ../../../Chr3

cp ../../../AI/Chr3/Present/pedlist1.dat .
cp ../../../AI/Chr3/Present/ai_pre* .
cp ../../../AI/Chr3/Absent/pedlist2.dat .
cp ../../../AI/Chr3/Absent/ai_abs* .

cd ../../../Chr5

cp ../../../AI/Chr5/Present/pedlist1.dat .
cp ../../../AI/Chr5/Present/ai_pre* .
cp ../../../AI/Chr5/Absent/pedlist2.dat .
cp ../../../AI/Chr5/Absent/ai_abs* .

cd ../../../Chr9

cp ../../../AI/Chr9/Present/pedlist1.dat .
cp ../../../AI/Chr9/Present/ai_pre* .
cp ../../../AI/Chr9/Absent/pedlist2.dat .
cp ../../../AI/Chr9/Absent/ai_abs* .
# Move into the next population DA and repeat the analyses.

cd DA
mkdir Chr1 Chr3 Chr5 Chr9
cd Chr1
mkdir Present Absent
cd Present

cp ../..../da.dat .
cp ../..../datain.dat .
cp ../..../map.dat .
awk '$3 > 0' da.dat > tmp1.dat
awk '$6 > 1' tmp1.dat > tmp2.dat
awk '$41 < 2' tmp2.dat > tmp3.dat
awk '$42 < 2' tmp3.dat > tmp4.dat
awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat
rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file da.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped ('cat pedlist.dat')
# awk \\
"$1 == 'ped'" da.dat >> pedin.dat

# end

# or (2) code to call pull_pedigrees.pl to do the work:

pull_pedigrees.pl da.dat pedlist1.dat pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_pre_01.lod
awk '/B01T0558/' merlin_out.01 > da_pre_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_pre_03.lod
awk '/B03T3056/' merlin_out.03 > da_pre_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_pre_05.lod
awk '/B05T4140/' merlin_out.05 > da_pre_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_pre_09.lod
awk '/B09T8333/' merlin_out.09 > da_pre_09.lod

cd ../Absent

cp .././././da.dat .
cp .././././datain.dat .
cp .././././map.dat .

# compare ai.dat with pedin.dat in the directory-present, then retain
# those
# lines that are not in pedin.dat and write it to a file called pedin.dat

cp ../Present/pedin.dat pedin1.dat
cat ./pedin1.dat ./da.dat | sort | uniq -u > pedin.dat
awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_abs_01.lod

awk '/B01T0558/' merlin_out.01 > da_abs_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_abs_03.lod
awk '/B03T3056/' merlin_out.03 > da_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_abs_05.lod
awk '/B05T4140/' merlin_out.05 > da_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_abs_09.lod
awk '/B09T8333/' merlin_out.09 > da_abs_09.lod

cd ../..

# Move into the next : Chr3/
cd Chr3

# Repeating all the steps done for Chr1
mkdir Present Absent
cd Present
  cp ../../*da.dat .
  cp ../../*datain.dat .
  cp ../../*map.dat .
  awk '$3 > 0' da.dat > tmp1.dat
  awk '$6 > 1' tmp1.dat > tmp2.dat
  awk '$117 < 2' tmp2.dat > tmp3.dat
  awk '$118 < 2' tmp3.dat > tmp4.dat
  awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat
rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file ai.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)
#   awk '$1 == '$ped'' da.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl da.dat pedlist1.dat pedin.dat

# running mega2

    mega2 << end_of_input
    2
dat
    0
    27
    y
    4
    1
    0
    8
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_pre_01.lod

awk '/B01T0558/' merlin_out.01 > da_pre_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_pre_03.lod

awk '/B03T3056/' merlin_out.03 > da_pre_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_pre_05.lod

awk '/B05T4140/' merlin_out.05 > da_pre_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_pre_09.lod

awk '/B09T8333/' merlin_out.09 > da_pre_09.lod
cd ../Absent

cp ../././da.dat .

cp ../././datain.dat .

cp ../././map.dat .

cp .././Present/pedin.dat pedin1.dat

cat ./pedin1.dat ./da.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_abs_01.lod
awk '/B01T0558/' merlin_out.01 > da_abs_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_abs_03.lod

awk '/B03T3056/' merlin_out.03 > da_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_abs_05.lod

awk '/B05T4140/' merlin_out.05 > da_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_abs_09.lod

awk '/B09T8333/' merlin_out.09 > da_abs_09.lod

cd ../..

# Move into the next: Chr5

cd Chr5

# Repeat steps

mkdir Present Absent

cd Present

91
cp ../../../../da.dat .
cp ../../../../datain.dat .
cp ../../../../map.dat .
awk 'S$3 > 0' da.dat > tmp1.dat
awk 'S$6 > 1' tmp1.dat > tmp2.dat
awk 'S$165 < 2' tmp2.dat > tmp3.dat
awk 'S$166 < 2' tmp3.dat > tmp4.dat
awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat
rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file da.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)  
# awk "$1 == "$ped" da.dat >> pedin.dat  
# end

# or (2) code to call pull_pedigrees.pl to do the work:

pull_pedigrees.pl da.dat pedlist1.dat pedin.dat

# running mega2
mega2 "" end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_pre_01.lod

awk '/B01T0558/' merlin_out.01 > da_pre_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_pre_03.lod

awk '/B03T3056/' merlin_out.03 > da_pre_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_pre_05.lod

awk '/B05T4140/' merlin_out.05 > da_pre_05.lod
awk '/B05T4140/' merlin_out.05 > da_pre_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_pre_09.lod

awk '/B09T8333/' merlin_out.09 > da_pre_09.lod

cd ../../../Absent
cp ../../../da.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
cp ../../../Present/pedin.dat pedin1.dat
cat ./pedin1.dat ./da.dat | sort | uniq -u > pedin.dat
awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

2
dat
0

27
y

4
1
0
8
0

--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_abs_01.lod

  awk '/B01T0558/' merlin_out.01 > da_abs_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_abs_03.lod

  awk '/B03T3056/' merlin_out.03 > da_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_abs_05.lod

  awk '/B05T4140/' merlin_out.05 > da_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_abs_09.lod

  awk '/B09T8333/' merlin_out.09 > da_abs_09.lod

cd ..../
# Move into the next : Chr9

`cd Chr9`

# Repeat steps

`mkdir Present Absent`
`cd Present`
`cp ../../../da.dat .`
`cp ../../../datain.dat .`
`cp ../../../map.dat .`
`awk '$3 > 0' da.dat > tmp1.dat`
`awk '$6 > 1' tmp1.dat > tmp2.dat`
`awk '$191 < 2' tmp2.dat > tmp3.dat`
`awk '$192 < 2' tmp3.dat > tmp4.dat`
`awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat`
`rm tmp*`

# if the number in the pedlist.dat matches the number in the first field
# of the file da.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`) 
# awk "$1 == "$ped"" da.dat >> pedin.dat 
# end
# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl da.dat pedlist1.dat pedin.dat

# running mega2

    mega2 << end_of_input

    2
dat
0
27
y
4
1
0
8
0

    --npl --markerNames --bits 38

    end_of_input

    ./merlin.all.sh

# finding the peaks in the output files in merlin:

    # print the lines that contain the pattern B01T0558 in
    # the file merlin_out.01 and write it to the file da_pre_01.lod

    awk '/B01T0558/' merlin_out.01 > da_pre_01.lod
# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_pre_03.lod

awk '/B03T3056/' merlin_out.03 > da_pre_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_pre_05.lod

awk '/B05T4140/' merlin_out.05 > da_pre_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_pre_09.lod

awk '/B09T8333/' merlin_out.09 > da_pre_09.lod

cd ../Absent

cp ../../../da.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
cp ../../../Present/pedin.dat pedin1.dat
cat ./pedin1.dat ./da.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

2
dat
0
#!/merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_abs_01.lod

awk '/B01T0558/' merlin_out.01 > da_abs_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_abs_03.lod

awk '/B03T3056/' merlin_out.03 > da_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_abs_05.lod

awk '/B05T4140/' merlin_out.05 > da_abs_05.lod
# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_abs_09.lod

awk '/B09T8333/' merlin_out.09 > da_abs_09.lod

# Move up by three levels

cd ../../..

# Move into the DA_results directory and copy the results only

cd DA_results
mkdir Chr1 Chr3 Chr5 Chr9
cd Chr1
cp ../../DA/Chr1/Present/pedlist1.dat .
cp ../../DA/Chr1/Present/da_pre* .
cp ../../DA/Chr1/Absent/pedlist2.dat .
cp ../../DA/Chr1/Absent/da_abs* .

cd ../Chr3
cp ../../DA/Chr3/Present/pedlist1.dat .
cp ../../DA/Chr3/Present/da_pre* .
cp ../../DA/Chr3/Absent/pedlist2.dat .
cp ../../DA/Chr3/Absent/da_abs* .

cd ../Chr5
cp ../../DA/Chr5/Present/pedlist1.dat .
cp ../../DA/Chr5/Present/da_pre*.
cp ../../DA/Chr5/Absent/pedlist2.dat.
cp ../../DA/Chr5/Absent/da_abs*.

cd ../Chr9

cp ../../DA/Chr9/Present/pedlist1.dat.
cp ../../DA/Chr9/Present/da_pre*.
cp ../../DA/Chr9/Absent/pedlist2.dat.
cp ../../DA/Chr9/Absent/da_abs*.

cd ../..

# Move into the next population KA and repeat the analyses.

cd KA

mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1

mkdir Present Absent

cd Present

cp .././.././ka.dat.
cp .././.././datain.dat.
cp .././.././map.dat.
awk '$3 > 0' ka.dat > tmp1.dat
awk '$6 > 1' tmp1.dat > tmp2.dat
awk '$41 < 2' tmp2.dat > tmp3.dat
awk '$42 < 2' tmp3.dat > tmp4.dat
awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat
rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file ka.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)  
# awk "$1 == "$ped"" ka.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl ka.dat pedlist1.dat pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
27
y
4
1
0

102
awk '/B01T0558/' merlin_out.01 > ka_pre_01.lod

awk '/B03T3056/' merlin_out.03 > ka_pre_03.lod

awk '/B05T4140/' merlin_out.05 > ka_pre_05.lod

awk '/B09T8333/' merlin_out.09 > ka_pre_09.lod
cd ../Absent

cp ../..../ka.dat .
cp ../..../datain.dat .
cp ../..../map.dat .

# compare ka.dat with pedin.dat in the directory-present, then retain those
# lines that are not in pedin.dat and write it to a file called pedin.dat

cp ../Present/pedin.dat pedin1.dat

cat ./pedin1.dat ./ka.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_abs_01.lod

    awk '/B01T0558/' merlin_out.01 > ka_abs_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_abs_03.lod

    awk '/B03T3056/' merlin_out.03 > ka_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_abs_05.lod

    awk '/B05T4140/' merlin_out.05 > ka_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_abs_09.lod

    awk '/B09T8333/' merlin_out.09 > ka_abs_09.lod

cd ../..
# Move into the next : Chr3/

    cd Chr3

# Repeating all the steps done for Chr1
mkdir Present Absent

cd Present

cp ../../../ka.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
awk '$3 > 0' ka.dat > tmp1.dat
awk '$6 > 1' tmp1.dat > tmp2.dat
awk '$117 < 2' tmp2.dat > tmp3.dat
awk '$118 < 2' tmp3.dat > tmp4.dat
awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat
rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file ka.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)      
#  awk "$1 == "$ped" ka.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl ka.dat pedlist1.dat pedin.dat

# running mega2
mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_pre_01.lod

awk '/B01T0558/' merlin_out.01 > ka_pre_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_pre_03.lod

awk '/B03T3056/' merlin_out.03 > ka_pre_03.lod
# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_pre_05.lod

awk '/B05T4140/' merlin_out.05 > ka_pre_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_pre_09.lod

awk '/B09T8333/' merlin_out.09 > ka_pre_09.lod

cd ../Absent
cp ../../../ka.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
cp ../../../Present/pedin.dat pedin1.dat
cat ./pedin1.dat ./ka.dat | sort | uniq -u > pedin.dat
awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2
mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_abs_01.lod

awk '/B01T0558/' merlin_out.01 > ka_abs_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_abs_03.lod

awk '/B03T3056/' merlin_out.03 > ka_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_abs_05.lod

awk '/B05T4140/' merlin_out.05 > ka_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_abs_09.lod

awk '/B09T8333/' merlin_out.09 > ka_abs_09.lod
cd ../..

# Move into the next: Chr5

cd Chr5

# Repeat steps

mkdir Present Absent

cd Present

    cp ../../../ka.dat .
    cp ../../../datain.dat .
    cp ../../../map.dat .
    awk '$3 > 0' ka.dat > tmp1.dat
    awk '$6 > 1' tmp1.dat > tmp2.dat
    awk '$165 < 2' tmp2.dat > tmp3.dat
    awk '$166 < 2' tmp3.dat > tmp4.dat
    awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat

rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file ka.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:

# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)
# awk '$1 == 'Sped'' ka.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

pull_pedigrees.pl ka.dat pedlist1.dat pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_pre_01.lod
awk '/B01T0558/' merlin_out.01 > ka_pre_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_pre_03.lod
awk '/B03T3056/' merlin_out.03 > ka_pre_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_pre_05.lod
awk '/B05T4140/' merlin_out.05 > ka_pre_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_pre_09.lod
awk '/B09T8333/' merlin_out.09 > ka_pre_09.lod

cd ../Absent
cp ../..../ka.dat .
cp ../..../datain.dat .
cp ../..../map.dat .
cp ../..../Present/pedin.dat pedin1.dat
cat ./pedin1.dat ./ka.dat | sort | uniq -u > pedin.dat
awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

2
awk '/B01T0558/' merlin_out.01 > ka_abs_01.lod

awk '/B03T3056/' merlin_out.03 > ka_abs_03.lod

awk '/B05T4140/' merlin_out.05 > ka_abs_05.lod
awk '/B05T4140/' merlin_out.05 > ka_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_abs_09.lod

awk '/B09T8333/' merlin_out.09 > ka_abs_09.lod

cd ../..

# Move into the next : Chr9

cd Chr9

# Repeat steps

mkdir Present Absent

cd Present

cp ../../../ka.dat .

cp ../../../datain.dat .

cp ../../../map.dat .

awk '$3 > 0' ka.dat > tmp1.dat

awk '$6 > 1' tmp1.dat > tmp2.dat

awk '$191 < 2' tmp2.dat > tmp3.dat

awk '$192 < 2' tmp3.dat > tmp4.dat

awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat

rm tmp*
# if the number in the pedlist.dat matches the number in the first field
# of the file ka.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)
#   awk '$1 == $ped' ka.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl ka.dat pedlist1.dat pedin.dat

# running mega2
mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_pre_01.lod

awk '/B01T0558/' merlin_out.01 > ka_pre_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_pre_03.lod

awk '/B03T3056/' merlin_out.03 > ka_pre_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_pre_05.lod

awk '/B05T4140/' merlin_out.05 > ka_pre_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_pre_09.lod

awk '/B09T8333/' merlin_out.09 > ka_pre_09.lod

cd ..../Absent

cp ../../../ka.dat .
cp ../../../../datain.dat .
cp ../../../../map.dat .
cp ../Present/pedin.dat pedin1.dat
cat ./pedin1.dat /ka.dat | sort | uniq -u > pedin.dat
   awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input
  2
dat
  0
  27
  y
  4
  1
  0
  8
  0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_abs_01.lod

awk '/B01T0558/' merlin_out.01 > ka_abs_01.lod
# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_abs_03.lod

awk '/B03T3056/' merlin_out.03 > ka_abs_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_abs_05.lod

awk '/B05T4140/' merlin_out.05 > ka_abs_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_abs_09.lod

awk '/B09T8333/' merlin_out.09 > ka_abs_09.lod

# Move up by three levels

cd ../../..

# Move into the KA_results directory and copy the results only

cd KA_results
mkdir Chr1 Chr3 Chr5 Chr9
cd Chr1
cp ../../KA/Chr1/Present/pedlist1.dat .
cp ../../KA/Chr1/Present/ka_pre*. 
cp ../../KA/Chr1/Absent/pedlist2.dat .

118
cp ../../KA/Chr1/Absent/ka_abs*. 

cd ../Chr3

cp ../../KA/Chr3/Present/pedlist1.dat .
cp ../../KA/Chr3/Present/ka_pre*. 
cp ../../KA/Chr3/Absent/pedlist2.dat .
cp ../../KA/Chr3/Absent/ka_abs*. 

cd ../Chr5

cp ../../KA/Chr5/Present/pedlist1.dat .
cp ../../KA/Chr5/Present/ka_pre*. 
cp ../../KA/Chr5/Absent/pedlist2.dat .
cp ../../KA/Chr5/Absent/ka_abs*. 

cd ../Chr9

cp ../../KA/Chr9/Present/pedlist1.dat .
cp ../../KA/Chr9/Present/ka_pre*. 
cp ../../KA/Chr9/Absent/pedlist2.dat .
cp ../../KA/Chr9/Absent/ka_abs*. 

cd ../..

# Move into NY and repeat the analyses (this time using simwalk2)

cd NY

mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1

mkdir Present Absent
cd Present

cp ../../../ny.dat .

cp ../../../datain.dat .

cp ../../../map.dat .

awk "$3 > 0" ny.dat > tmp1.dat

awk "$6 > 1" tmp1.dat > tmp2.dat

awk "$41 < 2" tmp2.dat > tmp3.dat

awk "$42 < 2" tmp3.dat > tmp4.dat

awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat

rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file ny.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)  
#   awk "$1 == '"ped"' ny.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl ny.dat pedlist1.dat pedin.dat

# running mega2

mega2 << end_of_input

120
mv map.dat map1.dat

./npl.all.sh

# copy the results into different files

cp STATS-01.ALL ny_pre_01.lod
cp STATS-03.ALL ny_pre_03.lod
cp STATS-05.ALL ny_pre_05.lod
cp STATS-09.ALL ny_pre_09.lod

cd ../Absent
cp ../../../ny.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
# compare ny.dat with pedin.dat in the directory-present, then retain those
# lines that are not in pedin.dat and write it to a file called pedin.dat

```bash
cp ../Present/pedin.dat pedin1.dat
cat ./pedin1.dat ./ny.dat | sort | uniq -u > pedin.dat
awk '{print $1}' pedin.dat | sort -u > pedlist2.dat
```

# running mega2

```bash
mega2 << end_of_input
2
dat
0
1
3
y
4
1
8
0
end_of_input
```

mv map.dat map1.dat

```bash
./npl.all.sh
```

# Copy results into files

```bash
cp STATS-01.ALL ny_abs_01.lod
cp STATS-03.ALL ny_abs_03.lod
cp STATS-05.ALL ny_abs_05.lod
cp STATS-09.ALL ny_abs_09.lod
```
cd ../../../

# Move into the next : Chr3/

cd Chr3

# Repeating all the steps done for Chr1

mkdir Present Absent

cd Present

cp ../../../ny.dat .

cp ../../../datain.dat .

cp ../../../map.dat .

awk '3 > 0' ny.dat > tmp1.dat

awk '6 > 1' tmp1.dat > tmp2.dat

awk '117 < 2' tmp2.dat > tmp3.dat

awk '118 < 2' tmp3.dat > tmp4.dat

awk '{print 1}' tmp4.dat | sort -u > pedlist1.dat

rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file ny.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:

# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`
#   awk '$1 == '$ped' ny.dat >> pedin.dat

awk '1 == '$ped' ny.dat >> pedin.dat

awk 'Sped' ny.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl ny.dat pedlist1.dat pedin.dat

# running mega2

    mega2 <<< end_of_input

    2
data
    0
    1
    3
    y
    4
    1
    8
    0

    end_of_input

    mv map.dat map1.dat

    ./npl.all.sh

# Copy results into files

    cp STATS-01.ALL ny_pre_01.lod
    cp STATS-03.ALL ny_pre_03.lod
cp STATS-05.ALL ny_pre_05.lod

cp STATS-09.ALL ny_pre_09.lod

cd ../Absent

cp ../../../ny.dat .
cp ../../../datain.dat .
cp ../../../map.dat .

cp ../../../Present/pedin.dat pedin1.dat
cat ../../../pedin1.dat ../../../ny.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

2
dat
0
1
3
y
4
1
8
0
end_of_input

mv map.dat map1.dat

./npl.all.sh
# Copy results into files

```
cp STATS-01.ALL ny_abs_01.lod

cp STATS-03.ALL ny_abs_03.lod

cp STATS-05.ALL ny_abs_05.lod

cp STATS-09.ALL ny_abs_09.lod
```

```
cd ../..
```

# Move into the next: Chr5

```
cd Chr5
```

# Repeat steps

```
mkdir Present Absent

cd Present

```

```
cp ../../../ny.dat .

cp ../../../datain.dat .

cp ../../../map.dat .

```

```
awk '$3 > 0' ny.dat > tmp1.dat

```

```
awk '$6 > 1' tmp1.dat > tmp2.dat

```

```
awk '$165 < 2' tmp2.dat > tmp3.dat

```

```
awk '$166 < 2' tmp3.dat > tmp4.dat

```

```
awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat

```

```
rm tmp*
```

# if the number in the pedlist.dat matches the number in the first field

# of the file ny.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)
#  awk "$1 == '$ped'" ny.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl ny.dat pedlist1.dat pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
1
3
y
4
1
8
0
end_of_input

mv map.dat map1.dat
# Copy results into files

cp STATS-01.ALL ny_pre_01.lod

cp STATS-03.ALL ny_pre_03.lod

cp STATS-05.ALL ny_pre_05.lod

cp STATS-09.ALL ny_pre_09.lod

cd ../Absent

cp ../../../ny.dat .

cp ../../../datain.dat .

cp ../../../map.dat .

cp ../../../Present/pedin.dat pedin1.dat

cat ./pedin1.dat ./ny.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

2
dat
0
1
3
y
4
1
8
mv map.dat map1.dat

./npl.all.sh

# Copy results into files
cp STATS-01.ALL ny_abs_01.lod
cp STATS-03.ALL ny_abs_03.lod
cp STATS-05.ALL ny_abs_05.lod
cp STATS-09.ALL ny_abs_09.lod

cd ../..

# Move into the next : Chr9

cd Chr9

# Repeat steps

mkdir Present Absent

cd Present

cp ../../../ny.dat .

cp ../../../datain.dat .

cp ../../../map.dat .

awk '$3 > 0' ny.dat > tmp1.dat

awk '$6 > 1' tmp1.dat > tmp2.dat
awk '$191 < 2' tmp2.dat > tmp3.dat
awk '$192 < 2' tmp3.dat > tmp4.dat
awk '{print $1}' tmp4.dat | sort -u > pedlist1.dat
rm tmp*

# if the number in the pedlist.dat matches the number in the first field
# of the file ny.dat, then retain those lines and write it to a
# file called pedin.dat

# two ways to do this:
# (1) code in this file to loop over pedigrees in pedlist.dat:

# foreach ped (`cat pedlist.dat`)
#   awk '$1 == "$ped"' ny.dat >> pedin.dat
# end

# or (2) code to call pull_pedigrees.pl to do the work:

    pull_pedigrees.pl ny.dat pedlist1.dat pedin.dat

# running mega2

mega2 << end_of_input
2

dat
0
1
3
y
mv map.dat map1.dat

./npl.all.sh

# Copy results into files

cp STATS-01.ALL ny_pre_01.lod

cp STATS-03.ALL ny_pre_03.lod

cp STATS-05.ALL ny_pre_05.lod

cp STATS-09.ALL ny_pre_09.lod

cd ../Absent

cp ../../../ny.dat .

cp ../../../datain.dat .

cp ../../../map.dat .

cp ../../../Present/pedin.dat pedin1.dat

cat ./pedin1.dat ./ny.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

# running mega2

mega2 << end_of_input

2
dat
0
mv map.dat map1.dat

./npl.all.sh

# Copy results into files
cp STATS-01.ALL ny_abs_01.lod
cp STATS-03.ALL ny_abs_03.lod
cp STATS-05.ALL ny_abs_05.lod
cp STATS-09.ALL ny_abs_09.lod

cd ../../..

# Move into the NY_results directory and copy the results only

cd NY_results
mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1

cp ../../../NY/Chr1/Present/pedlist1.dat .

cp ../../../NY/Chr1/Present/ny_pre* .

end_of_input
cp ../../NY/Chr1/Absent/pedlist2.dat .
cp ../../NY/Chr1/Absent/ny_abs* .

cd ../Chr3

cp ../..../NY/Chr3/Present/pedlist1.dat .
cp ../..../NY/Chr3/Present/ny_pre* .
cp ../..../NY/Chr3/Absent/pedlist2.dat .
cp ../..../NY/Chr3/Absent/ny_abs* .

cd ../Chr5

cp ../..../NY/Chr5/Present/pedlist1.dat .
cp ../..../NY/Chr5/Present/ny_pre* .
cp ../..../NY/Chr5/Absent/pedlist2.dat .
cp ../..../NY/Chr5/Absent/ny_abs* .

cd ../Chr9

cp ../..../NY/Chr9/Present/pedlist1.dat .
cp ../..../NY/Chr9/Present/ny_pre* .
cp ../..../NY/Chr9/Absent/pedlist2.dat .
cp ../..../NY/Chr9/Absent/ny_abs* .

cd ../..../

cd Redef

# REDEFINITION ANALYSES

mkdir AI DA KA NY
mkdir AI_results DA_results KA_results NY_results
cd AI
mkdir Chr1 Chr3 Chr5 Chr9
cd Chr1
mkdir Homoz Heteroz
cd Homoz

cp ../../../ai.dat .
cp ../../../datain.dat .
cp ../../../map.dat .

# redefining phenotypes: affecteds that are homozygous for the '1' allele
awk '{if($6 == 2 && $41 + $42 == 2) print $1, $2, $3, $4, $5, $6;
     else print $1, $2, $3, $4, $5, "1"}'} ai.dat > pheno.out
cut -f 7- ai.dat > geno.out
pastes pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_homoz_01.lod

awk '/B01T0558/" merlin_out.01 > ai_homoz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_homoz_03.lod

awk '/B03T3056/" merlin_out.03 > ai_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_homoz_05.lod

awk '/B05T4140/" merlin_out.05 > ai_homoz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_homoz_09.lod

awk '/B09T8333/" merlin_out.09 > ai_homoz_09.lod

cd ../Heteroz
cp ..../ai.dat .
cp ..../datain.dat .
# the next group of redefined phenotypes with the affecteds and
# heterozygotes as affecteds
awk '{if($6 == 2 && $41 + $42 == 3) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}'} ai.dat > pheno.out
cut -f7- ai.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2
mega2 <<< end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:
# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_heteroz_01.lod

awk '/B01T0558/' merlin_out.01 > ai_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_heteroz_03.lod

awk '/B03T3056/' merlin_out.03 > ai_heteroz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_heteroz_05.lod

awk '/B05T4140/' merlin_out.05 > ai_heteroz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_heteroz_09.lod

awk '/B09T8333/' merlin_out.09 > ai_heteroz_09.lod

cd ../../Chr3

mkdir Homoz Heteroz

cd Homoz

cp ../../../ai.dat .
cp ../../../datain.dat .
cp ../../../map.dat .

awk '{if($6 == 2 && $117 + $118 == 2) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}’ ai.dat > pheno.out

cut -f7 ai.dat > geno.out

paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2

dat

0

27

y

4

1

0

8

0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_homoz_01.lod

awk '/B01T0558/' merlin_out.01 > ai_homoz_01.lod

138
# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_homoz_03.lod

awk '/B03T3056/' merlin_out.03 > ai_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_homoz_05.lod

awk '/B05T4140/' merlin_out.05 > ai_homoz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_homoz_09.lod

awk '/B09T8333/' merlin_out.09 > ai_homoz_09.lod

cd ../Heteroz

cp ../../../ai.dat .
cp ../../../datain.dat .
cp ../../../map.dat .

# the next group of redefined phenotypes with the affecteds and
# heterozygotes as affecteds

awk '{if($6 == 2 && $117 + $118 == 3) print $1, $2, $3, $4, $5, $6;
   else print $1, $2, $3, $4, $5, "1"}’ ai.dat > pheno.out
cut -f7- ai.dat > geno.out
paste pheno.out geno.out > pedin.dat
# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_heteroz_01.lod

awk '/B01T0558/' merlin_out.01 > ai_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_heteroz_03.lod
awk '/B03T3056/" merlin_out.03 > ai_heteroz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_heteroz_05.lod

awk '/B05T4140/" merlin_out.05 > ai_heteroz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_heteroz_09.lod

awk '/B09T8333/" merlin_out.09 > ai_heteroz_09.lod

cd ../../Chr5
mkdir Homoz Heteroz
cd Homoz
cp ../../ai.dat .
cp ../../datain.dat .
cp ../../map.dat .
awk '{if($6 == 2 && $165 + $166 == 4) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}' ai.dat > pheno.out
cut -f7- ai.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
/merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_homoz_01.lod

awk '/B01T0558/' merlin_out.01 > ai_homoz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_homoz_03.lod

awk '/B03T3056/' merlin_out.03 > ai_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_homoz_05.lod

awk '/B05T4140/' merlin_out.05 > ai_homoz_05.lod
awk '/B09T8333/' merlin_out.09 > ai_homoz_09.lod

cd ../.Heteroz

cp ..../ai.dat .

cp ../../datain.dat .

cp ../../map.dat .

awk '{if($6 == 2 && $165 + $166 == 3) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}’ ai.dat > pheno.out

cut -f 7- ai.dat > geno.out

paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_heteroz_01.lod

awk '/B01T0558/' merlin_out.01 > ai_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_heteroz_03.lod

awk '/B03T3056/' merlin_out.03 > ai_heteroz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_heteroz_05.lod

awk '/B05T4140/' merlin_out.05 > ai_heteroz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_heteroz_09.lod
awk '/B09T8333/' merlin_out.09 > ai_heteroz_09.lod

cd ../../Chr9
mkdir Homoz Heteroz

cd Homoz

cp ../../../ai.dat .

cp ../../../datain.dat .

cp ../../../map.dat .

awk '{if($6 == 2 && $191 + $192 == 4) print $1, $2, $3, $4, $5, $6; else print $1, $2, $3, $4, $5, "1"}' ai.dat > pheno.out

cut -f7- ai.dat > geno.out

paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_homoz_01.lod
awk '/B01T0558/' merlin_out.01 > ai_homoz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_homoz_03.lod
awk '/B03T3056/' merlin_out.03 > ai_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_homoz_05.lod
awk '/B05T4140/' merlin_out.05 > ai_homoz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_homoz_09.lod
awk '/B09T8333/' merlin_out.09 > ai_homoz_09.lod

cd ../Heteroz
cp ../../../ai.dat .
cp ../../../datain.dat .
cp ../../../../map.dat .

# the next group of redefined phenotypes with the affecteds and
# heterozygotes as affecteds
awk '{if($6 == 2 && $191 + $192 == 3) print $1, $2, $3, $4, $5, $6;
     else print $1, $2, $3, $4, $5, "1"}' ai.dat > pheno.out

cut -f7- ai.dat > geno.out

paste pheno.out geno.out > pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:
# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_heteroz_01.lod

awk '/B01T0558/' merlin_out.01 > ai_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_heteroz_03.lod

awk '/B03T3056/' merlin_out.03 > ai_heteroz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_heteroz_05.lod

awk '/B05T4140/' merlin_out.05 > ai_heteroz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_heteroz_09.lod

awk '/B09T8333/' merlin_out.09 > ai_heteroz_09.lod

cd ../../../
# Move into the AI_results directory and copy the results only

cd AI_results
mkdir Chr1 Chr3 Chr5 Chr9
cd Chr1
cp ../../AI/Chr1/Homoz/ai_homoz* .
cp ../../AI/Chr1/Heteroz/ai_heteroz* .

cd ../Chr3

cp ../../AI/Chr3/Homoz/ai_homoz* .
cp ../../AI/Chr3/Heteroz/ai_heteroz* .

cd ../Chr5

cp ../../AI/Chr5/Homoz/ai_homoz* .
cp ../../AI/Chr5/Heteroz/ai_heteroz* .

cd ../Chr9

cp ../../AI/Chr9/Homoz/ai_homoz* .
cp ../../AI/Chr9/Heteroz/ai_heteroz* .

cd ../../

# Move into the next population DA and repeat the analyses.

cd DA
mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1
mkdir Homoz Heteroz

cd Homoz

cp ../../da.dat .
cp ../../datain.dat .
cp ../../map.dat .
awk '{if($6 == 2 && $41 + $42 == 2) print $1, $2, $3, $4, $5, $6;}'
else print $1, $2, $3, $4, $5, "1"}
d ata.out > pheno.out
cut -f7 da.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

 2
dat
 0
 27
  y
  4
  1
  0
  8
  0
--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_homoz_01.lod

awk '/B01T0558/' merlin_out.01 > da_homoz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_homoz_03.lod

awk '/B03T3056/' merlin_out.03 > da_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_homoz_05.lod

awk '/B05T4140/' merlin_out.05 > da_homoz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_homoz_09.lod

awk '/B09T8333/' merlin_out.09 > da_homoz_09.lod

cd ../Heteroz

cp ../../da.dat .
cp ../../datain.dat .
cp ../../map.dat .

# the next group of redefined phenotypes with the affecteds and
# heterozygotes as affecteds

awk '{if($6 == 2 && $41 + $42 == 3) print $1, $2, $3, $4, $5, $6;
 else print $1, $2, $3, $4, $5, "1")}' da.dat > pheno.out

cut -f 7- da.dat > geno.out

paste pheno.out geno.out > pedin.dat
# running mega2
mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_heteroz_01.lod

awk '/B01T0558/' merlin_out.01 > da_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_heteroz_03.lod

awk '/B03T3056/' merlin_out.03 > da_heteroz_03.lod
awk '/B05T4140/' merlin_out.05 > da_heteroz_05.lod

awk '/B09T8333/' merlin_out.09 > da_heteroz_09.lod

cd ../../Chr3

mkdir Homoz Heteroz

cd Homoz

cp ../../../da.dat .
cp ../../../datain.dat .
cp ../../../map.dat .

awk '{if($6 == 2 && $117 + $118 == 2) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}' da.dat > pheno.out

cut -f7- da.dat > geno.out

cpaste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_homoz_01.lod

awk '/B01T0558/' merlin_out.01 > da_homoz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_homoz_03.lod

awk '/B03T3056/' merlin_out.03 > da_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_homoz_05.lod

awk '/B05T4140/' merlin_out.05 > da_homoz_05.lod
# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_homoz_09.lod

awk '/B09T8333/' merlin_out.09 > da_homoz_09.lod

cd ../Heteroz
cp ../..../da.dat .
cp ../..../datain.dat .
cp ../..../map.dat .

# the next group of redefined phenotypes with the affecteds and
# heterozygotes as affecteds
awk '{if($6 == 2 && $117 + $118 == 3) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}'} da.dat > pheno.out

cut -f 7- da.dat > geno.out

copy pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
awk '/B01T0558/' merlin_out.01 > da_heteroz_01.lod
awk '/B03T3056/' merlin_out.03 > da_heteroz_03.lod
awk '/B05T4140/' merlin_out.05 > da_heteroz_05.lod
awk '/B09T8333/' merlin_out.09 > da_heteroz_09.lod
awk '/B09T8333/' merlin_out.09 > da_heteroz_09.lod

cd ../../Chr5
mkdir Homoz Heteroz

cd Homoz

    cp ../..../da.dat .

    cp ../..../datain.dat .

    cp ../..../map.dat .

    awk '{if($6 == 2 && $165 + $166 == 4) print $1, $2, $3, $4, $5, $6;

    else print $1, $2, $3, $4, $5, "1")' da.dat > pheno.out

    cut -f 7- da.dat > geno.out

    paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

derm_of_input
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_homoz_01.lod

awk '/B01T0558/' merlin_out.01 > da_homoz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_homoz_03.lod

awk '/B03T3056/' merlin_out.03 > da_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_homoz_05.lod

awk '/B05T4140/' merlin_out.05 > da_homoz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_homoz_09.lod

awk '/B09T8333/' merlin_out.09 > da_homoz_09.lod

cd ../Heteroz
cp ../..../da.dat .
cp ../..../datain.dat .
cp ../../../../map.dat .

# the next group of redefined phenotypes with the affecteds and
# heterozygotes as affecteds
awk '{if($6 == 2 && $165 + $166 == 3) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}’ da.dat > pheno.out
cut -f7- da.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:
# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_heteroz_01.lod
awk '/B01T0558/' merlin_out.01 > da_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_heteroz_03.lod
awk '/B03T3056/' merlin_out.03 > da_heteroz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_heteroz_05.lod
awk '/B05T4140/' merlin_out.05 > da_heteroz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_heteroz_09.lod
awk '/B09T8333/' merlin_out.09 > da_heteroz_09.lod

cd ../../Chr9
mkdir Homoz Heteroz
cd Homoz

cp ../../../da.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
awk '{if($6 == 2 && $191 + $192 == 4) print $1, $2, $3, $4, $5, $6;}'

160
else print $1, $2, $3, $4, $5, "$1")' da.dat > pheno.out
cut -f 7- da.dat > geno.out
paste pheno.out geno.out > pedin.dat
# running mega2
mega2 << end_of_input
2
    dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_homoz_01.lod

awk '/B01T0558/' merlin_out.01 > da_homoz_01.lod

# print the lines that contain the pattern B03T3056 in


# the file merlin_out.03 and write it to the file da_homoz_03.lod

awk '/B03T3056/' merlin_out.03 > da_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_homoz_05.lod

awk '/B05T4140/' merlin_out.05 > da_homoz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_homoz_09.lod

awk '/B09T8333/' merlin_out.09 > da_homoz_09.lod

cd ../Heteroz

cp .././../da.dat .

cp .././../datain.dat .

cp .././../map.dat .

# the next group of redefined phenotypes with the affecteds and
# heterozygotes as affecteds

awk '{if($6 == 2 && $191 + $192 == 3) print $1, $2, $3, $4, $5, $6;
     else print $1, $2, $3, $4, $5, "1")' da.dat > pheno.out

cut -f 7- da.dat > geno.out

paste pheno.out geno.out > pedin.dat
# running mega2

mega2 << end_of_input

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in 
# the file merlin_out.01 and write it to the file da_heteroz_01.lod

awk '/B01T0558/' merlin_out.01 > da_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in 
# the file merlin_out.03 and write it to the file da_heteroz_03.lod

awk '/B03T3056/' merlin_out.03 > da_heteroz_03.lod
# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_heteroz_05.lod

awk '/B05T4140/' merlin_out.05 > da_heteroz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_heteroz_09.lod

awk '/B09T8333/' merlin_out.09 > da_heteroz_09.lod

# Move into DA_results and copy the results

cd ../../../DA_results
mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1

cp ../../../DA/Chr1/Homoz/da_homoz* .
cp ../../../DA/Chr1/Heteroz/da_heteroz* .

cd ../Chr3

cp ../../../DA/Chr3/Homoz/da_homoz* .
cp ../../../DA/Chr3/Heteroz/da_heteroz* .

cd ../Chr5

cp ../../../DA/Chr5/Homoz/da_homoz* .
cp ../../../DA/Chr5/Heteroz/da_heteroz* .

164
cd ..../Chr9

cp ../../../DA/Chr9/Homoz/da_homoz*. 

cp ../../../DA/Chr9/Heteroz/da_heteroz*.

cd ../../../

# Move into KA and repeat the analyses

cd KA

mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1

mkdir Homoz Heteroz

cd Homoz

    cp ../../../ka.dat .

    cp ../../../datain.dat .

    cp ../../../map.dat .

    awk '{if($6 == 2 && $41 + $42 == 2) print $1, $2, $3, $4, $5, $6; 
        else print $1, $2, $3, $4, $5, "1"}'} ka.dat > pheno.out

    cut -f 7- ka.dat > geno.out

    paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
#!/merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_homoz_01.lod

awk '/B01T0558/' merlin_out.01 > ka_homoz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_homoz_03.lod

awk '/B03T3056/' merlin_out.03 > ka_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_homoz_05.lod

awk '/B05T4140/' merlin_out.05 > ka_homoz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_homoz_09.lod
awk '/B09T8333/' merlin_out.09 > ka_homoz_09.lod

cd ../Heteroz

cp ../../../ka.dat .
cp ../../../datain.dat .
cp ../../../map.dat .

# the next group of redefined phenotypes with the affecteds and
# heterozygotes as affecteds

awk '{if($6 == 2 && $41 + $42 == 3) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}'} ka.dat > pheno.out
cut -f 7- ka.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_heteroz_01.lod

awk '/B01T0558/' merlin_out.01 > ka_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_heteroz_03.lod

awk '/B03T3056/' merlin_out.03 > ka_heteroz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_heteroz_05.lod

awk '/B05T4140/' merlin_out.05 > ka_heteroz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_heteroz_09.lod

awk '/B09T8333/' merlin_out.09 > ka_heteroz_09.lod
cd ../../Chr3

mkdir Homoz Heteroz

cd Homoz

cp ../../../..//ka.dat .
cp ../../../datain.dat .
cp ../../../map.dat .

awk '{if($6 == 2 && $117 + $118 == 2) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}’ ka.dat > pheno.out
cut -f 7- ka.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_homoz_01.lod

awk '/B01T0558/' merlin_out.01 > ka_homoz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_homoz_03.lod

awk '/B03T3056/' merlin_out.03 > ka_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_homoz_05.lod

awk '/B05T4140/' merlin_out.05 > ka_homoz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_homoz_09.lod

awk '/B09T8333/' merlin_out.09 > ka_homoz_09.lod

cd ../Heteroz
cp ../../../ka.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
# the next group of redefined phenotypes with the affecteds and
# heterozygotes as affecteds

awk '{if($6 == 2 && $117 + $118 == 3) print $1, $2, $3, $4, $5, $6;\
     else print $1, $2, $3, $4, $5, "1")' ka.dat > pheno.out

cut -f7- ka.dat > geno.out

paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:
# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_heteroz_01.lod

awk '/B01T0558/' merlin_out.01 > ka_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_heteroz_03.lod

awk '/B03T3056/' merlin_out.03 > ka_heteroz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_heteroz_05.lod

awk '/B05T4140/' merlin_out.05 > ka_heteroz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_heteroz_09.lod

awk '/B09T8333/' merlin_out.09 > ka_heteroz_09.lod

cd ../../../Chr5
mkdir Homoz Heteroz

cd Homoz

cp ../../../ka.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
awk '{if($6 == 2 && $165 + $166 == 4) print $1, $2, $3, $4, $5, $6;
     else print $1, $2, $3, $4, $5, "1")}' ka.dat > pheno.out
cut -f 7- ka.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_homoz_01.lod
awk '/B01T0558/' merlin_out.01 > ka_homoz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_homoz_03.lod
awk '/'B03T3056/' merlin_out.03 > ka_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_homoz_05.lod
awk '/'B05T4140/' merlin_out.05 > ka_homoz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_homoz_09.lod
awk '/'B09T8333/' merlin_out.09 > ka_homoz_09.lod

cd ../Heteroz
cp ../../../ka.dat .
cp ../../../datain.dat .
cp ../../../map.dat .

# the next group of redefined phenotypes with the affecteds and
# heterozygotes as affecteds
awk '{if($6 == 2 && $165 + $166 == 3) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}'} ka.dat > pheno.out
cut -f 7- ka.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_heteroz_01.lod

awk '/B01T0558/' merlin_out.01 > ka_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_heteroz_03.lod

awk '/B03T3056/' merlin_out.03 > ka_heteroz_03.lod

175
# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_heteroz_05.lod

awk '/B05T4140/' merlin_out.05 > ka_heteroz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_heteroz_09.lod

awk '/B09T8333/' merlin_out.09 > ka_heteroz_09.lod

cd ../../Chr9
mkdir Homoz Heteroz

cd Homoz
cp ../../../ka.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
awk '{if($6 == 2 && $191 + $192 == 4) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}'} ka.dat > pheno.out
cut -f 7- ka.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input
2
dat
0
27
y
4
./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_homoz_01.lod

awk '/B01T0558/' merlin_out.01 > ka_homoz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_homoz_03.lod

awk '/B03T3056/' merlin_out.03 > ka_homoz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_homoz_05.lod

awk '/B05T4140/' merlin_out.05 > ka_homoz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_homoz_09.lod
awk '/B09T8333/' merlin_out.09 > ka_homoz_09.lod

cd ../Heteroz

cp ../..../ka.dat .

cp ../..../datain.dat .

cp ../..../map.dat .

# the next group of redefined phenotypes with the affecteds and heterozygotes as affecteds

awk '{if($6 == 2 && $191 + $192 == 3) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}' ka.dat > pheno.out

cut -f7- ka.dat > geno.out

paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_heteroz_01.lod

awk '/B01T0558/' merlin_out.01 > ka_heteroz_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_heteroz_03.lod

awk '/B03T3056/' merlin_out.03 > ka_heteroz_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_heteroz_05.lod

awk '/B05T4140/' merlin_out.05 > ka_heteroz_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_heteroz_09.lod

awk '/B09T8333/' merlin_out.09 > ka_heteroz_09.lod
cd ../../../

# Move into the KA_results directory and copy the results

cd KA_results
mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1

cp ../../KA/Chr1/Homoz/ka_homoz* .
cp ../../KA/Chr1/Heteroz/ka_heteroz* .

cd ../Chr3

cp ../../KA/Chr3/Homoz/ka_homoz* .
cp ../../KA/Chr3/Heteroz/ka_heteroz* .

cd ../Chr5

cp ../../KA/Chr5/Homoz/ka_homoz* .
cp ../../KA/Chr5/Heteroz/ka_heteroz* .

cd ../Chr9

cp ../../KA/Chr9/Homoz/ka_homoz* .
cp ../../KA/Chr9/Heteroz/ka_heteroz* .

cd ../..

# Move into NY and repeat the analyses (using simwalk2)

cd NY
mkdir Chr1 Chr3 Chr5 Chr9
cd Chr1
mkdir Homoz Heteroz
cd Homoz
  cp ../../../ny.dat .
  cp ../../../datain.dat .
  cp ../../../map.dat .
  awk '{if($6 == 2 && $41 + $42 == 2) print $1, $2, $3, $4, $5, $6;
    else print $1, $2, $3, $4, $5, "1"}'} ai.dat > pheno.out
cut -f 7- ai.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
1
3
y
4
1
8
0
end_of_input

mv map.dat map1.dat

./npl.all.sh
# Copy results into files

cp STATS-01.ALL ny_homoz_01.lod

cp STATS-03.ALL ny_homoz_03.lod

cp STATS-05.ALL ny_homoz_05.lod

cp STATS-09.ALL ny_homoz_09.lod

cd ../Heteroz

cp ../../../ny.dat .

cp ../../../datain.dat .

cp ../../../map.dat .

awk '{if($6 == 2 && $41 + $42 == 3) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1")}' ai.dat > pheno.out

cut -f 7- ai.dat > geno.out

paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat

0

1

3

y

4

1

8
mv map.dat map1.dat

./npl.all.sh

# Copy results into files
cp STATS-01.ALL ny_heteroz_01.lod
cp STATS-03.ALL ny_heteroz_03.lod
cp STATS-05.ALL ny_heteroz_05.lod
cp STATS-09.ALL ny_heteroz_09.lod

cd ../..
# Move into the next : Chr3/
cd Chr3

# Repeating all the steps done for Chr1

mkdir Homoz Heteroz
cd Homoz

cp ./../../ny.dat .
cp ./../../datain.dat .
cp ./../../map.dat .
awk '{if($6 == 2 && $117 + $118 == 2) print $1, $2, $3, $4, $5, $6;
   else print $1, $2, $3, $4, $5, "1")} ai.dat > pheno.out
cut -f7 ai.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
1
3
y
4
1
8
0
end_of_input

mv map.dat map1.dat

./npl.all.sh

# Copy results into files
cp STATS-01.ALL ny_homoz_01.lod
cp STATS-03.ALL ny_homoz_03.lod
cp STATS-05.ALL ny_homoz_05.lod
cp STATS-09.ALL ny_homoz_09.lod

cd ../Heteroz
cp ../../../ny.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
awk '{if($6 == 2 && $117 + $118 == 3) print $1, $2, $3, $4, $5, $6;
     else print $1, $2, $3, $4, $5, "1"}'} ai.dat > pheno.out
cut -f 7- ai.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input
2
dat
0
1
3
y
4
1
8
0
end_of_input

mv map.dat map1.dat

../npl.all.sh

# Copy results into files
cp STATS-01.ALL ny_heteroz_01.lod
cp STATS-03.ALL ny_heteroz_03.lod
cp STATS-05.ALL ny_heteroz_05.lod
cp STATS-09.ALL ny_heteroz_09.lod

cd ../..

# Move into the next: Chr5

cd Chr5

# Repeat steps

mkdir Homoz Heteroz

cd Homoz

cp ../../ny.dat .
cp ../../datain.dat .
cp ../../map.dat .

awk '{if($6 == 2 && $165 + $166 == 4) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1")' ai.dat > pheno.out

cut -f7- ai.dat > geno.out

cpaste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input
2

dat
mv map.dat map1.dat

./npl.all.sh

# Copy results into files
cp STATS-01.ALL ny_homoz_01.lod
cp STATS-03.ALL ny_homoz_03.lod
cp STATS-05.ALL ny_homoz_05.lod
cp STATS-09.ALL ny_homoz_09.lod

cd ../Heteroz

    cp ../../../ny.dat .
    cp ../../../datain.dat .
    cp ../../../map.dat .

    awk '{if($6 == 2 && $165 + $166 == 3) print $1, $2, $3, $4, $5, $6;
        else print $1, $2, $3, $4, $5, "1"}’ ai.dat > pheno.out
    cut -f7- ai.dat > geno.out
    paste pheno.out geno.out > pedin.dat
# running mega2

mega2 << end_of_input

2
dat
0
1
3
y
4
1
8
0
end_of_input

mv map.dat map1.dat

./npl.all.sh

# Copy results into files

cp STATS-01.ALL ny_heteroz_01.lod

cp STATS-03.ALL ny_heteroz_03.lod

cp STATS-05.ALL ny_heteroz_05.lod

cp STATS-09.ALL ny_heteroz_09.lod

cd ../..
# Move into the next : Chr9

cd Chr9

# Repeat steps

mkdir Homoz Heteroz

cd Homoz

cp ../../../ny.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
awk '{if($6 == 2 && $191 + $192 == 4) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}’ ai.dat > pheno.out
cut -f 7- ai.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2

mega2 << end_of_input

2
dat
0
1
3
y
4
1
8
mv map.dat map1.dat

./npl.all.sh

# Copy results into files
cp STATS-01.ALL ny_homoz_01.lod
cp STATS-03.ALL ny_homoz_03.lod
cp STATS-05.ALL ny_homoz_05.lod
cp STATS-09.ALL ny_homoz_09.lod

cd ../Heteroz

cp ../../../ny.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
awk '{if($6 == 2 && $191 + $192 == 3) print $1, $2, $3, $4, $5, $6;
else print $1, $2, $3, $4, $5, "1"}’ ai.dat > pheno.out
cut -f7- ai.dat > geno.out
paste pheno.out geno.out > pedin.dat

# running mega2
mega2 << end_of_input
2
dat
0
mv map.dat map1.dat

./npl.all.sh

# Copy results into files
cp STATS-01.ALL ny_heteroz_01.lod
cp STATS-03.ALL ny_heteroz_03.lod
cp STATS-05.ALL ny_heteroz_05.lod
cp STATS-09.ALL ny_heteroz_09.lod

cd ../../../

# Move into NY_results and copy the results

cd NY_results
mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1

cp ../../../NY/Chr1/Homoz/ny_homoz* .

cp ../../NY/Chr1/Heteroz/ny_heteroz* .

cd ../Chr3

cp ../..//NY/Chr3/Homoz/ny_homoz* .
cp ../..//NY/Chr3/Heteroz/ny_heteroz* .

cd ../Chr5

cp ../..//NY/Chr5/Homoz/ny_homoz* .
cp ../..//NY/Chr5/Heteroz/ny_heteroz* .

cd ../Chr9

cp ../..//NY/Chr9/Homoz/ny_homoz* .
cp ../..//NY/Chr9/Heteroz/ny_heteroz* .

cd .././.

# Weighted analyses

cd Weight

mkdir AI DA KA NY
mkdir AI_results DA_results KA_results NY_results

# AI population - weighted analyses

cd AI

mkdir Chr1 Chr3 Chr5 Chr9 Genomescan

cd Genomescan

cp .././ai.dat .
cp .././datain.dat .
cp ../../../map.dat.
cp ai.dat pedin.dat

# running mega2
mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38 --perFamily
dat

end_of_input

# tweak the merlin shell scripts to add the line about merlin.lod

# extension changes

    echo "mv merlin.lod merlin.01.lod">>merlin.01.sh
    echo "mv merlin.lod merlin.03.lod">>merlin.03.sh
    echo "mv merlin.lod merlin.05.lod">>merlin.05.sh
    echo "mv merlin.lod merlin.09.lod">>merlin.09.sh

./merlin.all.sh


cd ../Chr1

mkdir Pos Neg

cd Pos
cp ../../Genomescan/merlin.01.lod.

# print the lines with the pattern 'B01T0558'
awk '/B01T0558/' merlin.01.lod > tmp1

# take merlin.01.lod and retain only the fields 1 & 8
awk '{print $1, $8}' tmp1 > tmp2

# retain lines if column 2 > 0 ---> tmp3
awk '$2>0' tmp2 > tmp3

# print the first field (pedigree numbers) to pedlist.dat
awk '{print $1}' tmp3 > pedlist1.dat

# choose the pedigrees based on pedlist.dat from ai.dat
cp ../../../ai.dat .

cp ../../../datain.dat .

cp ../../../map.dat .

pull_pedigrees.pl ai.dat pedlist1.dat pedin.dat

mega2 << end_of_input

2
data
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

dat
end_of_input
/merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_pos_01.lod

awk '/B01T0558/' merlin_out.01 > ai_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_pos_03.lod

awk '/B03T3056/' merlin_out.03 > ai_pos_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_pos_05.lod

awk '/B05T4140/' merlin_out.05 > ai_pos_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_pos_09.lod

awk '/B09T8333/' merlin_out.09 > ai_pos_09.lod

cd ../Neg

cp ../../../ai.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
cp ../../../Pos/pedin.dat pedin1.dat
cat ./pedin1.dat ./ai.dat | sort | uniq -u > pedin.dat
awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_neg_01.lod

awk '/B01T0558/' merlin_out.01 > ai_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_neg_03.lod

awk '/B03T3056/' merlin_out.03 > ai_neg_03.lod
# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_neg_05.lod

awk '/B05T4140/' merlin_out.05 > ai_neg_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_neg_09.lod

awk '/B09T8333/' merlin_out.09 > ai_neg_09.lod

cd ../../Chr3
mkdir Pos Neg
cd Pos
cp ../../Genomescan/merlin.03.lod .

# print the lines with the pattern 'B03T3056'
awk '/B03T3056/' merlin.03.lod > tmp1

# take merlin.03.lod and retain only the fields 1 & 8
awk '{print $1, $8}' tmp1 > tmp2

# retain lines if column 2 > 0 ---> tmp3
awk '$2>0' tmp2 > tmp3

# print the first field (pedigree numbers) to pedlist.dat
awk '{print $1}' tmp3 > pedlist1.dat

# choose the pedigrees based on pedlist1.dat from ai.dat

cp ../../ai.dat .
cp ../../datain.dat .
cp ../../map.dat .
pull_pedigrees.pl ai.dat pedlist1.dat pedin.dat
mega2 << end_of_input

12
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_pos_01.lod

    awk '/B01T0558/' merlin_out.01 > ai_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_pos_03.lod

    awk '/B03T3056/' merlin_out.03 > ai_pos_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_pos_05.lod

```bash
awk '/B05T4140/' merlin_out.05 > ai_pos_05.lod
```

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_pos_09.lod

```bash
awk '/B09T8333/' merlin_out.09 > ai_pos_09.lod
```

cd ../Neg

```bash
cp ../..../ai.dat .
cp ../..../datain.dat .
cp ../..../map.dat .
cp ../Pos/pedin.dat pedin1.dat
cat ./pedin1.dat ./ai.dat | sort | uniq -u > pedin.dat
    awk '{print $1}' pedin.dat | sort -u > pedlist2.dat
```

```bash
mega2 <<< end_of_input
2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
```
# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_neg_01.lod

awk '/B01T0558/' merlin_out.01 > ai_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_neg_03.lod

awk '/B03T3056/' merlin_out.03 > ai_neg_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_neg_05.lod

awk '/B05T4140/' merlin_out.05 > ai_neg_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_neg_09.lod

awk '/B09T8333/' merlin_out.09 > ai_neg_09.lod

cd ../../Chr5
mkdir Pos Neg

cd Pos

200
cp ../../Genomescan/merlin.05.lod .

# print the lines with the pattern 'B05T4140'
awk '/B05T4140/' merlin.05.lod > tmp1

# take merlin.01.lod and retain only the fields 1 & 8
awk '{print $1, $8}' tmp1 > tmp2

# retain lines if column 2 > 0 ---> tmp3
awk '$2>0' tmp2 > tmp3

# print the first field (pedigree numbers) to pedlist.dat
awk '{print $1}' tmp3 > pedlist1.dat

# choose the pedigrees based on pedlist1.dat from ai.dat

cp ../../../ai.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
pull_pedigrees.pl ai.dat pedlist1.dat pedin.dat

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input
./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_pos_01.lod

awk '/B01T0558/' merlin_out.01 > ai_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_pos_03.lod

awk '/B03T3056/' merlin_out.03 > ai_pos_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_pos_05.lod

awk '/B05T4140/' merlin_out.05 > ai_pos_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_pos_09.lod

awk '/B09T8333/' merlin_out.09 > ai_pos_09.lod

cd ../Neg

cp ../../*.dat .
cp ../../*.datain.dat .
cp ../../*.map.dat .
cp ../Pos/pedin.dat pedin1.dat
cat ./pedin1.dat ./ai.dat | sort | uniq -u > pedin.dat
awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_neg_01.lod

awk '/B01T0558/' merlin_out.01 > ai_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_neg_03.lod

awk '/B03T3056/' merlin_out.03 > ai_neg_03.lod

203
# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_neg_05.lod

awk '/B05T4140/' merlin_out.05 > ai_neg_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_neg_09.lod

awk '/B09T8333/' merlin_out.09 > ai_neg_09.lod

cd ../../Chr9
mkdir Pos Neg
cd Pos
cp ../../Genomescan/merlin.09.lod .

# print the lines with the pattern 'B09T8333'
awk '/B09T8333/' merlin.09.lod > tmp1

# take merlin.01.lod and retain only the fields 1 & 8
awk '{print $1, $8}' tmp1 > tmp2

# retain lines if column 2 > 0 ---> tmp3
awk '$2>0' tmp2 > tmp3

# print the first field (pedigree numbers) to pedlist.dat
awk '{print $1}' tmp3 > pedlist1.dat

# choose the pedigrees based on pedlist1.dat from ai.dat

cp ../../ai.dat .
cp ../../datain.dat .
cp ../../map.dat .
pull_pedigrees.pl ai.dat pedlist1.dat pedin.dat
./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_pos_01.lod

awk '/B01T0558/" merlin_out.01 > ai_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_pos_03.lod

awk '/B03T3056/" merlin_out.03 > ai_pos_03.lod

# print the lines that contain the pattern B05T4140 in

205
# the file merlin_out.05 and write it to the file ai_pos_05.lod

awk '/B05T4140/' merlin_out.05 > ai_pos_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_pos_09.lod

awk '/B09T8333/' merlin_out.09 > ai_pos_09.lod

cd ../Neg

cp ../..../ai.dat .
cp ../..../datain.dat .
cp ../..../map.dat .
cp ../Pos/pedin.dat pedin1.dat
cat ./pedin1.dat ./ai.dat | sort | uniq -u > pedin.dat
    awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

mega2 << end_of_input

2

dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ai_neg_01.lod

awk '/B01T0558/' merlin_out.01 > ai_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ai_neg_03.lod

awk '/B03T3056/' merlin_out.03 > ai_neg_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ai_neg_05.lod

awk '/B05T4140/' merlin_out.05 > ai_neg_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ai_neg_09.lod

awk '/B09T8333/' merlin_out.09 > ai_neg_09.lod

cd ../../../

# Move into AI_results and copy the files
cd AI_results
mkdir Chr1 Chr3 Chr5 Chr9
cd Chr1
cp ../AI/Chr1/Pos/pedlist1.dat .
cp ../AI/Chr1/Pos/ai_pos* .
cp ../AI/Chr1/Neg/pedlist2.dat .
cp ../AI/Chr1/Neg/ai_neg* .

cd ../Chr3
cp ../AI/Chr3/Pos/pedlist1.dat .
cp ../AI/Chr3/Pos/ai_pos* .
cp ../AI/Chr3/Neg/pedlist2.dat .
cp ../AI/Chr3/Neg/ai_neg* .

cd ../Chr5
cp ../AI/Chr5/Pos/pedlist1.dat .
cp ../AI/Chr5/Pos/ai_pos* .
cp ../AI/Chr5/Neg/pedlist2.dat .
cp ../AI/Chr5/Neg/ai_neg* .

cd ../Chr9
cp ../AI/Chr9/Pos/pedlist1.dat .
cp ../AI/Chr9/Pos/ai_pos* .
cp ../AI/Chr9/Neg/pedlist2.dat .
cp ../AI/Chr9/Neg/ai_neg* .

cd ../../
# Move into the next population DA and repeat the analyses.

cd DA
 mkdir Chr1 Chr3 Chr5 Chr9 Genomescan
cd Genomescan
    cp ../../../da.dat .
    cp ../../../datain.dat .
    cp ../../../map.dat .
    cp da.dat pedin.dat
# running mega2
mega2 << end_of_input
  2
dat
  0
  27
  y
  4
  1
  0
  8
  0
  --npl --markerNames --bits 38 --perFamily
end_of_input

# tweak the merlin shell scripts to add the line about merlin.lod
# extension changes
    echo "mv merlin.lod merlin.01.lod">>merlin.01.sh
echo "mv merlin.lod merlin.03.lod">>merlin.03.sh
echo "mv merlin.lod merlin.05.lod">>merlin.05.sh
echo "mv merlin.lod merlin.09.lod">>merlin.09.sh

./merlin.all.sh

./Chr1

mkdir Pos Neg
cd Pos
cp ../../../Genomescan/merlin.01.lod .

# print the lines with the pattern 'B01T0558'
awk '/B01T0558/' merlin.01.lod > tmp1

# take merlin.01.lod and retain only the fields 1 & 8
awk '{print $1, $8}' tmp1 > tmp2

# retain lines if column 2 > 0 ---> tmp3
awk '$2>0' tmp2 > tmp3

# print the first field (pedigree numbers) to pedlist.dat
awk '{print $1}' tmp3 > pedlist1.dat

# choose the pedigrees based on pedlist1.dat from da.dat
cp ../../../da.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
pull_pedigrees.pl da.dat pedlist1.dat pedin.dat

mega2 << end_of_input
2
dat
0
27
y
```bash
./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_pos_01.lod

awk '/B01T0558/' merlin_out.01 > da_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_pos_03.lod

awk '/B03T3056/' merlin_out.03 > da_pos_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_pos_05.lod

awk '/B05T4140/' merlin_out.05 > da_pos_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_pos_09.lod
```
awk '/B09T8333/' merlin_out.09 > da_pos_09.lod

cd ../Neg

cp ../../../da.dat .
cp ../../../datain.dat .
cp ../../../map.dat .
cp ../Pos/pedin.dat pedin1.dat
cat ./pedin1.dat ./da.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

cmpa2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in

212
# the file merlin_out.01 and write it to the file da_neg_01.lod

    awk '/B01T0558/' merlin_out.01 > da_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_neg_03.lod

    awk '/B03T3056/' merlin_out.03 > da_neg_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_neg_05.lod

    awk '/B05T4140/' merlin_out.05 > da_neg_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_neg_09.lod

    awk '/B09T8333/' merlin_out.09 > da_neg_09.lod

cd ../../Chr3
mkdir Pos Neg
cd Pos
cp ../../Genomescan/merlin.03.lod .
# print the lines with the pattern 'B03T3056'
    awk '/B03T3056/' merlin.03.lod > tmp1
# take merlin.03.lod and retain only the fields 1 & 8
    awk '{print $1, $8}' tmp1 > tmp2
# retain lines if column 2 > 0 --> tmp3
awk 'S2>0' tmp2 > tmp3

# print the first field (pedigree numbers) to pedlist.dat
awk '{print $1}' tmp3 > pedlist1.dat

# choose the pedigrees based on pedlist1.dat from da.dat

cp ../../../../da.dat .
cp ../../../../datain.dat .
cp ../../../../map.dat .
pull_pedigrees.pl da.dat pedlist1.dat pedin.dat

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_pos_01.lod

214
awk '/B01T0558/' merlin_out.01 > da_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_pos_03.lod

awk '/B03T3056/' merlin_out.03 > da_pos_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_pos_05.lod

awk '/B05T4140/' merlin_out.05 > da_pos_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_pos_09.lod

awk '/B09T8333/' merlin_out.09 > da_pos_09.lod

cd ../Neg

cp ../../da.dat .
cp ../../datain.dat .
cp ../../map.dat .
cp ../Pos/pedin.dat pedin1.dat
cat ./pedin1.dat ./da.dat | sort | uniq -u > pedin.dat
    awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

mega2 << end_of_input

2
dat
0
npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_neg_01.lod

awk '/B01T0558/' merlin_out.01 > da_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_neg_03.lod

awk '/B03T3056/' merlin_out.03 > da_neg_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_neg_05.lod

awk '/B05T4140/' merlin_out.05 > da_neg_05.lod
# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_neg_09.lod

    awk '/B09T8333/' merlin_out.09 > da_neg_09.lod

cd ../../Chr5
mkdir Pos Neg
cd Pos
    cp ../../Genomescan/merlin.05.lod .
# print the lines with the pattern 'B05T4140'
    awk '/B05T4140/' merlin.05.lod > tmp1
# take merlin.01.lod and retain only the fields 1 & 8
    awk '{print $1, $8}' tmp1 > tmp2
# retain lines if column 2 > 0 ---> tmp3
    awk '$2>0' tmp2 > tmp3
# print the first field (pedigree numbers) to pedlist.dat
    awk '{print $1}' tmp3 > pedlist1.dat
# choose the pedigrees based on pedlist1.dat from da.dat
    cp ../../da.dat .
    cp ../../datain.dat .
    cp ../../map.dat .
pull_pedigrees.pl da.dat pedlist1.dat pedin.dat
mega2 << end_of_input

2
dat
0
27
y
./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_pos_01.lod

awk '/B01T0558/' merlin_out.01 > da_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_pos_03.lod

awk '/B03T3056/' merlin_out.03 > da_pos_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_pos_05.lod

awk '/B05T4140/' merlin_out.05 > da_pos_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_pos_09.lod

awk '/B09T8333/' merlin_out.09 > da_pos_09.lod

218
awk '/B09T8333/' merlin_out.09 > da_pos_09.lod

cd ../Neg
cp .././././da.dat .
cp .././././datain.dat .
cp .././././map.dat .
cp ../Pos/pedin.dat pedin1.dat
cat ./pedin1.dat ./da.dat | sort | uniq -u > pedin.dat
    awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_neg_01.lod

```bash
awk '/B01T0558/' merlin_out.01 > da_neg_01.lod
```

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_neg_03.lod

```bash
awk '/B03T3056/' merlin_out.03 > da_neg_03.lod
```

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_neg_05.lod

```bash
awk '/B05T4140/' merlin_out.05 > da_neg_05.lod
```

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_neg_09.lod

```bash
awk '/B09T8333/' merlin_out.09 > da_neg_09.lod
```

cd ../../Chr9

```bash
mkdir Pos Neg
```

```bash
cd Pos
```

```bash
cp ../../Genomescan/merlin.09.lod .
```

# print the lines with the pattern 'B09T8333'

```bash
awk '/B09T8333/' merlin.09.lod > tmp1
```

# take merlin.01.lod and retain only the fields 1 & 8

```bash
awk '{print $1, $8}' tmp1 > tmp2
```

# retain lines if column 2 > 0 ---> tmp3

```bash
# retain lines if column 2 > 0 ---> tmp3
```
awk 'S2>0' tmp2 > tmp3

# print the first field (pedigree numbers) to pedlist.dat
awk '{print $1}' tmp3 > pedlist1.dat

# choose the pedigrees based on pedlist1.dat from da.dat
cp ../../../../da.dat .
cp ../../../../datain.dat .
cp ../../../../map.dat .
pull_pedigrees.pl da.dat pedlist1.dat pedin.dat

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_pos_01.lod
awk '/B01T0558/' merlin_out.01 > da_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_pos_03.lod
awk '/B03T3056/' merlin_out.03 > da_pos_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_pos_05.lod
awk '/B05T4140/' merlin_out.05 > da_pos_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file da_pos_09.lod
awk '/B09T8333/' merlin_out.09 > da_pos_09.lod

cd ../Neg

cp ../..../da.dat .
cp ../..../datain.dat .
cp ../..../map.dat .
cp ../Pos/pedin.dat pedin1.dat
cat ./pedin1.dat ./da.dat | sort | uniq -u > pedin.dat
    awk '{print $1}' pedin.dat | sort -u > pedlist2.dat
mega2 << end_of_input
2
dat
0
/merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file da_neg_01.lod

awk '/B01T0558/' merlin_out.01 > da_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file da_neg_03.lod

awk '/B03T3056/' merlin_out.03 > da_neg_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file da_neg_05.lod

awk '/B05T4140/' merlin_out.05 > da_neg_05.lod
awk '/B09T8333/' merlin_out.09 > da_neg_09.lod

cd ../../../

# Move into DA_results and copy the files

cd DA_results
mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1

cp ../DA/Chr1/Pos/pedlist1.dat .
cp ../DA/Chr1/Pos/da_pos* .
cp ../DA/Chr1/Neg/pedlist2.dat .
cp ../DA/Chr1/Neg/da_neg* .

cd ../Chr3

cp ../DA/Chr3/Pos/pedlist1.dat .
cp ../DA/Chr3/Pos/da_pos* .
cp ../DA/Chr3/Neg/pedlist2.dat .
cp ../DA/Chr3/Neg/da_neg* .

cd ../Chr5

cp ../DA/Chr5/Pos/pedlist1.dat .
cp ../DA/Chr5/Pos/da_pos* .
cp ../DA/Chr5/Neg/pedlist2.dat .
cp ../DA/Chr5/Neg/da_neg* .
cd ../Chr9

cp ../..//DA/Chr9/Pos/pedlist1.dat .
cp ../..//DA/Chr9/Pos/da_pos* .
cp ../..//DA/Chr9/Neg/pedlist2.dat .
cp ../..//DA/Chr9/Neg/da_neg* .

cd ../..

# Move into the next population KA and repeat the analyses.

cd KA

mkdir Chr1 Chr3 Chr5 Chr9 Genomescan
cd Genomescan

cp ../../../ka.dat .
cp ../../../datain.dat .
cp ../../../map.dat .

cp ka.dat pedin.dat

# running mega2

mega2 << end_of_input

2

dat

0

27

y

4

1

0

225
# tweak the merlin shell scripts to add the line about merlin.lod
#
# extension changes

echo "mv merlin.lod merlin.01.lod">>merlin.01.sh

echo "mv merlin.lod merlin.03.lod">>merlin.03.sh

echo "mv merlin.lod merlin.05.lod">>merlin.05.sh

echo "mv merlin.lod merlin.09.lod">>merlin.09.sh

./merlin.all.sh

cd ..../Chr1

mkdir Pos Neg

cd Pos

cp ..../Genomescan/merlin.01.lod .

# print the lines with the pattern 'B01T0558'

awk '/B01T0558/' merlin.01.lod > tmp1

# take merlin.01.lod and retain only the fields 1 & 8

awk '{print $1, $8}' tmp1 > tmp2

# retain lines if column 2 > 0 ---> tmp3

awk '$2>0' tmp2 > tmp3

# print the first field (pedigree numbers) to pedlist.dat

awk '{print $1}' tmp3 > pedlist1.dat

# choose the pedigrees based on pedlist1.dat from ka.dat

cp .././././ka.dat .
cp .././././datain.dat .
cp ../../../map.dat .

pull_pedigrees.pl ka.dat pedlist1.dat pedin.dat

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_pos_01.lod

awk '/B01T0558/' merlin_out.01 > ka_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_pos_03.lod

awk '/B03T3056/' merlin_out.03 > ka_pos_03.lod

227
# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_pos_05.lod

awk '/B05T4140/' merlin_out.05 > ka_pos_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_pos_09.lod

awk '/B09T8333/' merlin_out.09 > ka_pos_09.lod

cd ../Neg

cp ../../ka.dat .
cp ../../datain.dat .
cp ../../map.dat .
cp ../Pos/pedin.dat pedin1.dat
cat ./pedin1.dat ./ka.dat | sort | uniq -u > pedin.dat
   awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
--npl --markerNames --bits 38
dermend_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_neg_01.lod

awk '/B01T0558/' merlin_out.01 > ka_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_neg_03.lod

awk '/B03T3056/' merlin_out.03 > ka_neg_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_neg_05.lod

awk '/B05T4140/' merlin_out.05 > ka_neg_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_neg_09.lod

awk '/B09T8333/' merlin_out.09 > ka_neg_09.lod

cd ../../Chr3
mkdir Pos Neg

cd Pos

cp ../../../Genomescan/merlin.03.lod .

# print the lines with the pattern 'B03T3056'
awk '/B03T3056/' merlin.03.lod > tmp1

# take merlin.03.lod and retain only the fields 1 & 8
awk '{print $1, $8}' tmp1 > tmp2

# retain lines if column 2 > 0 ---> tmp3
awk '$2>0' tmp2 > tmp3

# print the first field (pedigree numbers) to pedlist.dat
awk '{print $1}' tmp3 > pedlist1.dat

# choose the pedigrees based on pedlist1.dat from ka.dat

  cp ../../../ka.dat .
  cp ../../../datain.dat .
  cp ../../../map.dat .

  pull_pedigrees.pl ka.dat pedlist1.dat pedin.dat

mega2 << end_of_input

  2
dat
  0
  27
  y
  4
  1
  0
  8
  0

  --npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_pos_01.lod

    awk '/B01T0558/' merlin_out.01 > ka_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_pos_03.lod

    awk '/B03T3056/' merlin_out.03 > ka_pos_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_pos_05.lod

    awk '/B05T4140/' merlin_out.05 > ka_pos_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_pos_09.lod

    awk '/B09T8333/' merlin_out.09 > ka_pos_09.lod

cd ../Neg
   cp ../..../ka.dat .
   cp ../..../datain.dat .
cp ../../map.dat .

cp ../Pos/pedin.dat pedin1.dat

cat ./pedin1.dat ./ka.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

mega2 << end_of_input

2

dat

0

27

y

4

1

0

8

0

--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_neg_01.lod

awk '/B01T0558/' merlin_out.01 > ka_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_neg_03.lod

awk '/B03T3056/' merlin_out.03 > ka_neg_03.lod

232
awk '/B03T3056/' merlin_out.03 > ka_neg_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_neg_05.lod

awk '/B05T4140/' merlin_out.05 > ka_neg_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_neg_09.lod

awk '/B09T8333/' merlin_out.09 > ka_neg_09.lod

cd ../../Chr5
mkdir Pos Neg
cd Pos

cp ../../../Genomescan/merlin.05.lod .

# print the lines with the pattern 'B05T4140'
awk '/B05T4140/' merlin.05.lod > tmp1

# take merlin.01.lod and retain only the fields 1 & 8
awk '{print $1, $8}' tmp1 > tmp2

# retain lines if column 2 > 0 ---> tmp3
awk '$2>0' tmp2 > tmp3

# print the first field (pedigree numbers) to pedlist.dat
awk '{print $1}' tmp3 > pedlist1.dat

# choose the pedigrees based on pedlist1.dat from ka.dat

cp ../../../ka.dat .
cp ../../../datain.dat .
cp ../../../../map.dat .
pull_pedigrees.pl ka.dat pedlist1.dat pedin.dat

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

--npl --markerNames --bits 38
end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_pos_01.lod

awk '/B01T0558/' merlin_out.01 > ka_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_pos_03.lod

awk '/B03T3056/' merlin_out.03 > ka_pos_03.lod

234
awk '/B05T4140/' merlin_out.05 > ka_pos_05.lod

awk '/B09T8333/' merlin_out.09 > ka_pos_09.lod

cd ../Neg

cp ../..../ka.dat .
cp ../..../datain.dat .
cp ../..../map.dat .
cp ../Pos/pedin.dat pedin1.dat

cat ./pedin1.dat ./ka.dat | sort | uniq -u > pedin.dat

awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

mega2 << end_of_input
2
dat
0
27
y
4
1
0
8
--npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_neg_01.lod

  awk '/B01T0558/' merlin_out.01 > ka_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_neg_03.lod

  awk '/B03T3056/' merlin_out.03 > ka_neg_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_neg_05.lod

  awk '/B05T4140/' merlin_out.05 > ka_neg_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_neg_09.lod

  awk '/B09T8333/' merlin_out.09 > ka_neg_09.lod

cd ../../Chr9
mkdir Pos Neg

cd Pos

cp ../../Genomescan/merlin.09.lod .
# print the lines with the pattern 'B09T8333'
awk '/B09T8333/' merlin.09.lod > tmp1
# take merlin.01.lod and retain only the fields 1 & 8
awk '{print $1, $8}' tmp1 > tmp2
# retain lines if column 2 > 0 ---> tmp3
awk '$2>0' tmp2 > tmp3
# print the first field (pedigree numbers) to pedlist.dat
awk '{print $1}' tmp3 > pedlist1.dat
# choose the pedigrees based on pedlist1.dat from ka.dat

cp ../../../../ka.dat .
cp ../../../../datain.dat .
cp ../../../../map.dat .
pull_pedigrees.pl ka.dat pedlist1.dat pedin.dat

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0
--npl --markerNames --bits 38
./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_pos_01.lod

awk '/B01T0558/' merlin_out.01 > ka_pos_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_pos_03.lod

awk '/B03T3056/' merlin_out.03 > ka_pos_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_pos_05.lod

awk '/B05T4140/' merlin_out.05 > ka_pos_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_pos_09.lod

awk '/B09T8333/' merlin_out.09 > ka_pos_09.lod

cd ../Neg

cp ../../../ka.dat .
cp ../../../datain.dat .
cp ../../map.dat .

cp ../../../map.dat .

cp ../Pos/pedin.dat pedin1.dat

cat ./pedin1.dat ./ka.dat | sort | uniq -u > pedin.dat

   awk '{print $1}' pedin.dat | sort -u > pedlist2.dat

mega2 << end_of_input

2
dat
0
27
y
4
1
0
8
0

   --npl --markerNames --bits 38

end_of_input

./merlin.all.sh

# finding the peaks in the output files in merlin:

# print the lines that contain the pattern B01T0558 in
# the file merlin_out.01 and write it to the file ka_neg_01.lod

   awk '/B01T0558/' merlin_out.01 > ka_neg_01.lod

# print the lines that contain the pattern B03T3056 in
# the file merlin_out.03 and write it to the file ka_neg_03.lod

   awk '/B03T3056/' merlin_out.03 > ka_neg_03.lod

239
awk '/B03T3056/' merlin_out.03 > ka_neg_03.lod

# print the lines that contain the pattern B05T4140 in
# the file merlin_out.05 and write it to the file ka_neg_05.lod

awk '/B05T4140/' merlin_out.05 > ka_neg_05.lod

# print the lines that contain the pattern B09T8333 in
# the file merlin_out.09 and write it to the file ka_neg_09.lod

awk '/B09T8333/' merlin_out.09 > ka_neg_09.lod

cd ../../../

# Move into KA_results and copy the files/results

cd KA_results
mkdir Chr1 Chr3 Chr5 Chr9

cd Chr1

cp ../../../KA/Chr1/Pos/pedlist1.dat .
cp ../../../KA/Chr1/Pos/ka_pos* .
cp ../../../KA/Chr1/Neg/pedlist2.dat .
cp ../../../KA/Chr1/Neg/ka_neg* .

cd ../Chr3

cp ../../../KA/Chr3/Pos/pedlist1.dat .
cp ../../../KA/Chr3/Pos/ka_pos* .
cp ../../KA/Chr3/Neg/pedlist2.dat .
cp ../../KA/Chr3/Neg/ka_neg* .

cd ../Chr5

cp ../..//KA/Chr5/Pos/pedlist1.dat .
cp ../..//KA/Chr5/Pos/ka_pos* .
cp ../..//KA/Chr5/Neg/pedlist2.dat .
cp ../..//KA/Chr5/Neg/ka_neg* .

cd .././Chr9

cp ../..//KA/Chr9/Pos/pedlist1.dat .
cp ../..//KA/Chr9/Pos/ka_pos* .
cp ../..//KA/Chr9/Neg/pedlist2.dat .
cp ../..//KA/Chr9/Neg/ka_neg* .

cd ./.

# Move into the next population NY and repeat the analyses

cd NY

mkdir Chr1 Chr3 Chr5 Chr9 Genomescan
cd Genomescan

cp ../../ny.dat .
cp ../../datain.dat .
cp ../../map.dat .

cp ny.dat pedin.dat

# running mega2

mega2 << end_of_input
mv map.dat map1.dat

./npl.all.sh

# Copy results into files (these have to be analyzed in order to split the families)

cp STATS-01.ALL ny_01.lod

cp STATS-03.ALL ny_03.lod

cp STATS-05.ALL ny_05.lod

cp STATS-09.ALL ny_09.lod

# Move into NY_results and copy the files

cd ../../../NY_results

cp ../../../NY/Genomescan/ny_0* .

cd ../../../
APPENDIX B

SHELL SCRIPT FOR LOGISTIC REGRESSION ANALYSIS

#!/bin/tcsh -f # C-shell file name: logist.sh
#
#----------------------------------------------
#   Input file names:
#   ai.dat
#   da.dat
#   ka.dat
#   ny.dat
#----------------------------------------------
#
#mkdir AI DA KA NY
#mkdir AI_results DA_results KA_results NY_results
#cd AI
#mkdir Chr1_3 Chr1_5 Chr1_9 Chr3_5 Chr3_9 Chr5_9
#cd Chr1_3
    cp ../../../ai.dat .
    echo "1 200" > ai_1_3.dat
    awk '$3 == 0' ai.dat | awk '{print $6, $41, $42, $117, $118}' >> ai_1_3.dat
    lrmodel ai_1_3.dat ai_1_3.r1
cd ../Chr1_5
    cp .././ai.dat .
    echo "1 200" > ai_1_5.dat
    awk '$3 == 0' ai.dat | awk '{print $6, $41, $42, $165, $166}' >> ai_1_5.dat
    lrmodel ai_1_5.dat ai_1_5.r1

cd ../Chr1_9
    cp .././ai.dat .
    echo "1 200" > ai_1_9.dat
    awk '$3 == 0' ai.dat | awk '{print $6, $41, $42, $191, $192}' >> ai_1_9.dat
    lrmodel ai_1_9.dat ai_1_9.r1

cd ../Chr3_5
    cp .././ai.dat .
    echo "1 200" > ai_3_5.dat
    awk '$3 == 0' ai.dat | awk '{print $6, $117, $118, $165, $166}' >> ai_3_5.dat
    lrmodel ai_3_5.dat ai_3_5.r1

cd ../Chr3_9
    cp .././ai.dat .
    echo "1 200" > ai_3_9.dat
    awk '$3 == 0' ai.dat | awk '{print $6, $117, $118, $191, $192}' >> ai_3_9.dat
    lrmodel ai_3_9.dat ai_3_9.r1

cd ../Chr5_9
    cp .././ai.dat .
    echo "1 200" > ai_5_9.dat
    awk '$3 == 0' ai.dat | awk '{print $6, $165, $166, $191, $192}' >> ai_5_9.dat
    lrmodel ai_5_9.dat ai_5_9.r1

# Move into AI_results and copy the files
cd ../../AI_results

cp ../AI/Chr1_3/ai_1_3.r1 .
cp ../AI/Chr1_5/ai_1_5.r1 .
cp ../AI/Chr1_9/ai_1_9.r1 .
cp ../AI/Chr3_5/ai_3_5.r1 .
cp ../AI/Chr3_9/ai_3_9.r1 .
cp ../AI/Chr5_9/ai_5_9.r1 .

cd ../DA

mkdir Chr1_3 Chr1_5 Chr1_9 Chr3_5 Chr3_9 Chr5_9

cd Chr1_3

cp ../../da.dat .

echo "1 200" > da_1_3.dat

awk '$3 == 0' da.dat | awk '{print $6, $41, $42, $117, $118}'>> da_1_3.dat

lrmodel da_1_3.dat da_1_3.r1

cd ../Chr1_5

cp ../../da.dat .

echo "1 200" > da_1_5.dat

awk '$3 == 0' da.dat | awk '{print $6, $41, $42, $165, $166}'>> da_1_5.dat

lrmodel da_1_5.dat da_1_5.r1

cd ../Chr1_9

cp ../../da.dat .

echo "1 200" > da_1_9.dat

awk '$3 == 0' da.dat | awk '{print $6, $41, $42, $191, $192}'>> da_1_9.dat

lrmodel da_1_9.dat da_1_9.r1

cd ../Chr3_5

cp ../../da.dat .

245
echo "1 200" > da_3_5.dat
awk '$3 == 0' da.dat | awk '{print $6, $117, $118, $165, $166}' >> da_3_5.dat
lrmodel da_3_5.dat da_3_5.r1

cd ../Chr3_9

    cp ../../../da.dat .
    echo "1 200" > da_3_9.dat
    awk '$3 == 0' da.dat | awk '{print $6, $117, $118, $191, $192}' >> da_3_9.dat
    lrmodel da_3_9.dat da_3_9.r1

cd ../Chr5_9

    cp ../../../da.dat .
    echo "1 200" > da_5_9.dat
    awk '$3 == 0' da.dat | awk '{print $6, $165, $166, $191, $192}' >> da_5_9.dat
    lrmodel da_5_9.dat da_5_9.r1

# Move into DA_results and copy the files

cd ../../../DA_results

    cp ../DA/Chr1_3/da_1_3.r1 .
    cp ../DA/Chr1_5/da_1_5.r1 .
    cp ../DA/Chr1_9/da_1_9.r1 .
    cp ../DA/Chr3_5/da_3_5.r1 .
    cp ../DA/Chr3_9/da_3_9.r1 .
    cp ../DA/Chr5_9/da_5_9.r1 .

cd ../KA

    mkdir Chr1_3 Chr1_5 Chr1_9 Chr3_5 Chr3_9 Chr5_9

cd Chr1_3

    cp ../../../ka.dat .

246
echo "1 200" > ka_1_3.dat
awk '$3 == 0' ka.dat | awk '{print $6, $41, $42, $117, $118}' >> ka_1_3.dat
lrmodel ka_1_3.dat ka_1_3.r1

cd ../Chr1_5
    cp ../ka.dat .
    echo "1 200" > ka_1_5.dat
    awk '$3 == 0' ka.dat | awk '{print $6, $41, $42, $165, $166}' >> ka_1_5.dat
    lrmodel ka_1_5.dat ka_1_5.r1

cd ../Chr1_9
    cp ../ka.dat .
    echo "1 200" > ka_1_9.dat
    awk '$3 == 0' ka.dat | awk '{print $6, $41, $42, $191, $192}' >> ka_1_9.dat
    lrmodel ka_1_9.dat ka_1_9.r1

cd ../Chr3_5
    cp ../ka.dat .
    echo "1 200" > ka_3_5.dat
    awk '$3 == 0' ka.dat | awk '{print $6, $117, $118, $165, $166}' >> ka_3_5.dat
    lrmodel ka_3_5.dat ka_3_5.r1

cd ../Chr3_9
    cp ../ka.dat .
    echo "1 200" > ka_3_9.dat
    awk '$3 == 0' ka.dat | awk '{print $6, $117, $118, $191, $192}' >> ka_3_9.dat
    lrmodel ka_3_9.dat ka_3_9.r1

cd ../Chr5_9
    cp ../ka.dat .
    echo "1 200" > ka_5_9.dat
    awk '$3 == 0' ka.dat | awk '{print $6, $165, $166, $191, $192}' >> ka_5_9.dat
    lrmodel ka_5_9.dat ka_5_9.r1
# Move into KA_results and copy the files

cd ../../KA_results

cp ../KA/Chr1_3/ka_1_3.r1 .
cp ../KA/Chr1_5/ka_1_5.r1 .
cp ../KA/Chr1_9/ka_1_9.r1 .
cp ../KA/Chr3_5/ka_3_5.r1 .
cp ../KA/Chr3_9/ka_3_9.r1 .
cp ../KA/Chr5_9/ka_5_9.r1 .

cd ../NY

mkdir Chr1_3 Chr1_5 Chr1_9 Chr3_5 Chr3_9 Chr5_9

cd Chr1_3

cp ../ny.dat .

echo "1 100" > ny_1_3.dat

awk '$3 == 0' ny.dat | awk '{print $6, $41, $42, $117, $118}' >> ny_1_3.dat

lrmodel ny_1_3.dat ny_1_3.r1

cd ../Chr1_5

cp ../ny.dat .

echo "1 100" > ny_1_5.dat

awk '$3 == 0' ny.dat | awk '{print $6, $41, $42, $165, $166}' >> ny_1_5.dat

lrmodel ny_1_5.dat ny_1_5.r1

cd ../Chr1_9

cp ../ny.dat .

echo "1 100" > ny_1_9.dat

awk '$3 == 0' ny.dat | awk '{print $6, $41, $42, $191, $192}' >> ny_1_9.dat

lrmodel ny_1_9.dat ny_1_9.r1
cd ../Chr3_5
  cp ../../ny.dat .
  echo "1 100" > ny_3_5.dat
  awk '$3 == 0' ny.dat | awk '{print $6, $117, $118, $165, $166}' >> ny_3_5.dat
  lrmodel ny_3_5.dat ny_3_5.r1

cd ../Chr3_9
  cp ../../ny.dat .
  echo "1 100" > ny_3_9.dat
  awk '$3 == 0' ny.dat | awk '{print $6, $117, $118, $191, $192}' >> ny_3_9.dat
  lrmodel ny_3_9.dat ny_3_9.r1

cd ../Chr5_9
  cp ../../ny.dat .
  echo "1 100" > ny_5_9.dat
  awk '$3 == 0' ny.dat | awk '{print $6, $165, $166, $191, $192}' >> ny_5_9.dat
  lrmodel ny_5_9.dat ny_5_9.r1

# Move into NY_results and copy the files

cd ../../NY_results
  cp ../NY/Chr1_3/ny_1_3.r1 .
  cp ../NY/Chr1_5/ny_1_5.r1 .
  cp ../NY/Chr1_9/ny_1_9.r1 .
  cp ../NY/Chr3_5/ny_3_5.r1 .
  cp ../NY/Chr3_9/ny_3_9.r1 .
  cp ../NY/Chr5_9/ny_5_9.r1 .

cd ..


