MOLECULAR PHYSIOLOGY AND BIOPHYSICS

INVESTIGATING THE GENETIC SUSCEPTIBILITY TO MULTIPLE SCLEROSIS: A GENOMIC CONVERGENCE APPROACH

SHANNON JO KENEALY

Dissertation under the direction of Professor Jonathan L. Haines

Multiple sclerosis (MS) is a debilitating neuroimmunological and neuro-degenerative disease. Despite substantial evidence for polygenic inheritance, the MHC is the only region that clearly and consistently demonstrates linkage and association in MS studies. The goal of the work presented in this dissertation was to identify additional chromosomal regions harboring MS susceptibility genes. Our studies entailed a new genomic convergence approach incorporating information gained from positional (linkage and association) and functional (comparative sequence) studies. In conjunction with high-throughput genotyping and powerful new statistical analyses methods, this approach identified several regions suggesting the presence of MS loci.

We began our investigation with a genomic linkage screen that identified seven chromosomal regions of interest in a data set of multiplex MS families. To narrow these regions, we developed an approach for more detailed linkage studies that capitalized on new methods for rapid and accurate genotyping of SNPs. In addition to increasing marker coverage in each region, we genotyped an expanded data set and devised covariate analyses schemes to account for genetic effect in the MHC. This method

continued to provide evidence of linkage to several chromosomal regions and was successful in substantially narrowing two regions to only a few Mb.

We then developed a systematic approach to expedite follow-up association studies in the positional candidate regions. In an attempt to increase the likelihood of detecting variants associated with MS, we employed a novel method to select SNPs located in multi-species conserved sequences. Use of this method on chromosome 1q44 resulted in the identification of four subregions demonstrating significant association with MS susceptibility.

The work presented in this dissertation confirmed several regions warranting further investigation for genes conferring susceptibility to MS, including chromosomes 1q44, 2q35, 9q34, and 18p11. It is our hope that these studies will result in the discovery of several genes associated with MS and that our genomic convergence approach will provide researchers with a method for unraveling the genetic heterogeneity of MS and other complex genetic diseases.

Approved_____Date_____

INVESTIGATING THE GENETIC SUSCEPTIBILITY TO MULTIPLE SCLEROSIS: A GENOMIC CONVERGENCE APPROACH

By

Shannon Jo Kenealy

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Molecular Physiology and Biophysics

May, 2006

Nashville, Tennessee

Approved:

Professor Jonathan L. Haines

Professor Marshall L. Summar

Professor Douglas P. Mortlock

Professor John A. Phillips

Professor Subramaniam Sriram

Professor Scott M. Williams

To my parents, M. Douglas and Carol Kenealy,

for fostering curiosity about science.

To my perfect partner in every way, Steven,

for your endless support.

ACKNOWLEDGEMENTS

The work presented in this dissertation would not have been possible without the generous participation of families and the collaborating clinics and clinicians that referred patients to our studies. This work was supported in part by the following grants: National Multiple Sclerosis Society RG 2899 and RG2901, NIH NS32380 and NS26799, INSERM and the Ministry of Research (CRB), FRM (Action 2000), ARSEP, the Nancy Davis Centers Without Walls Foundation, and the Vanderbilt-Meharry Joint Training Grant in Human Genetics.

I would like to express my gratitude to the following:

Ph.D. Mentor: Dr. Jonathan L. Haines. I am extremely fortunate to have found an advisor who conducts science with ingenuity and integrity. I am forever indebted to Dr. Haines for his time, patience, and trust in me.

Dissertation Committee: Dr. Marshall L. Summar (Committee Chair), Dr. Douglas P. Mortlock, Dr. John A. Phillips, Dr. Subramaniam Sriram, and Dr. Scott M. Williams. I have immense respect for each of my committee members as scientists and as people. With their knowledge and passion, the field of disease gene discovery is in good hands.

Members of the Haines Lab (Present and Former): Dr. Nathalie Schnetz-Boutaud, Melissa J. Allen, Brent M. Anderson, Shana A. Crabtree, Erin K. Hennessey, Holli B.

Hutcheson, Xueying (Sharon) Liang, Kylee M. Spencer, and Krista C. Stanton. I thank all the members of the Haines lab for their scientific collaboration and for the friendships that made the journey enjoyable.

A huge thank you also goes to Lindsey A. Herrel for her substantial contribution to the MS project. Lindsey is an excellent scientist and will become an excellent physician. I have no doubt that she will make a difference in this world.

Members of the Center for Human Genetics Research (CHGR): Family Ascertainment Core, DNA Resources Core, Data Analysis Core, and Bioinformatics Core.

I would especially like to express my gratitude to Data Analysis Core member Yuki Bradford. I greatly appreciate Yuki's patience in training me to perform statistical analyses and respect her commitment to generate quality data and analyses in all aspects of our studies.

A huge thank you also goes to Lana M. Olson, my diehard triathlon buddy. Our ridiculous training schedule somehow helped me remain sane during a very insane time in my life. I'll miss having someone to chat with about sore muscles and the addiction to venti non-fat lattes.

Collaborators:

The Multiple Sclerosis Genetics Group (MSGG):

Department of Neurology, University of California San Francisco

Division of Epidemiology, University of California Berkeley

Center for Human Genetics, Duke University Medical Center

Center for Human Genetics Research, Vanderbilt University Medical Center

The International Multiple Sclerosis Genetics Consortium (IMSGC):

Neurology Unit, University of Cambridge (UK)

Center for Human Genetics, Duke University Medical Center

Harvard Medical School

The Broad Institute, Massachusetts Institute of Technology

Center for Neurologic Diseases, Brigham and Women's Hospital

Division of Epidemiology, University of California Berkeley

Department of Neurology, University of California San Francisco

Center for Human Genetics Research, Vanderbilt University Medical Center

TABLE OF CONTENTS

LIST OF TABLES

Appendix C Table Page

Appendix D Table Page

Appendix E Table Page

LIST OF FIGURES

Appendix C Figure Page

LIST OF ABBREVIATIONS / SYMBOLS

ASP	affected sibling pair
BLAST	Basic Local Alignment Searching Tool
bp	base pair
cDNA	complementary deoxyribonucleic acid
CEPH	Centre d'Etude du Polymorphisme Humain
CHGR	(Vanderbilt) Center for Human Genetics Research
cM	centiMorgan
CNS	central nervous system
CSF	cerebrospinal fluid
ºC	degrees Celsius
dbSNP	(NCBI) database of single nucleotide polymorphisms
DNA	deoxyribonucleic acid
dNTP	2'-deoxynucleoside 5'-triphosphate
DSP	discordant sibling pair
DZ	dizygotic (fraternal twins)
EAE	experimental autoimmune encephalomyelitis (mouse model)
EDTA	ethylenediaminetetraacetic acid
F primer	forward (PCR) primer
FBAT	family based association test
GAMES	Genetic Analysis of Multiple Sclerosis in Europeans
GASP	Genometric Analysis Simulation Program

GDB	(Human) Genome Database
hCV	human Celera variation
HLOD	two-point heterogeneity LOD score
htSNP	haplotype tagging SNP
HWE	Hardy-Weinberg equilibrium
H_2O	water
IBD	identical by descent
IBS	identical by state
IDDM	insulin dependent diabetes mellitus (“type 1 diabetes”)
IMSGC	International Multiple Sclerosis Genetics Consortium
kb	kilobase (pair)
LD	linkage disequilibrium
LOD	$base_{10}$ logarithm of the likelihood of odds (ratio for linkage)
MAF	minor allele frequency
Mb	megabase (pair)
MCS	multi-species conserved sequences
MgCl	magnesium chloride
mL	milliliter
MLS	maximum multipoint LOD score
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
MS	multiple sclerosis
MSGG	Multiple Sclerosis Genetics Group

MZ	monozygotic (identical twins)
μg	microgram
μL	microliter
μM	micromolar
NCBI	National Center for Biotechnology Information
ng	nanogram
NHGRI	National Human Genome Research Institute
NPL	nonparametric linkage
OMIM	Online Mendelian Inheritance in Man
OSA	ordered subset analysis
PBAT	power study for family based association test
PCR	polymerase chain reaction
PDT	pedigree disequilibrium test
R primer	reverse (PCR) primer
RA	rheumatoid arthritis
RNA	ribonucleic acid
rs / RefSeq	(NCBI) reference sequence
SAGE	serial analysis of gene expression
sib-pair	sibling pair
sib-TDT	sibling transmission disequilibrium test
SLE	systemic lupus erythematosus
SNP	single nucleotide polymorphism
T_A	annealing temperature

Taq	*Thermus aquaticus* (polymerase)
TDT	transmission disequilibrium test
TE	Tris EDTA buffer
TEMED	tetramethylethylenediamine
T_M	melting temperature
Tris	Tris [hydroxymethyl] aminomethane
UTR	untranslated region
VISTA	Visualization Tools for Alignment

Nucleotide Abbreviations:

A	adenine
C	cytosine
G	guanine
T	thymine

Gene Abbreviations:

APOC2	apolipoprotein C2
APOE	apolipoprotein E
FMN2	formin 2
HLA	human leukocyte antigen
ILT6	immunoglobulin-like transcript 6

MBP	myelin basic protein
MHC	major histocompatibility complex
OPCML	opioid binding protein / cell adhesion molecule-like
OPN3	opsin 3
PDRC	protein related to DAC and cerberus
PVRL2	poliovirus receptor related protein 2
RGS7	regulator of G protein signaling 7
RYR2	ryanodine receptor 2
WDR64	WD repeat domain 64

Statistical Abbreviations / Symbols:

λ	recurrence risk ratio
λ_s	sibling recurrence risk ratio

LIST OF WEBSITES

ABI	http://www.appliedbiosystems.com/
Affymetrix	http://www.affymetrix.com/index.affx
BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
Celera	http://www.celeradiscoverysystem.com/index.cfm
CHGR (Internet)	http://chgr.mc.vanderbilt.edu/chgrsite/
CHGR (Supplemental Data)	http://chgr.mc.vanderbilt.edu/publications.html
decode	http://www.decode.com/
Ensembl	http://www.ensembl.org
GASP	http://research.nhgri.nih.gov/gasp/
GDB	http://www.gdb.org/
Haploview	http://www.broad.mit.edu/mpg/haploview/
HapMap	http://www.hapmap.org/
Illumina	http://www.illumina.com
Marshfield	http://research.marshfieldclinic.org/genetics/Map_Markers/maps/IndexMapFrames.html
NCBI	http://www.ncbi.nlm.nih.gov/
NCBI dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP
NHGRI WebMCS	http://research.nhgri.nih.gov/MCS/
OMIM	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM
Operon	http://www.operon.com/
Parallele Biosciences	http://www.parallelebio.com

Primer3	http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi
RepeatMasker	http://www.repeatmasker.org/
UCSC	http://genome.ucsc.edu/

CHAPTER I

INTRODUCTION*

Clinical Aspects of MS

Multiple sclerosis (MS) is an autoimmune inflammatory disease characterized by demyelination and neurodegeneration within the central nervous system (CNS) (MS [MIM 126200]). As the name of the disease implies, affected individuals exhibit hardened (or "sclerotic") tissue in many (or "multiple") parts of the brain and spinal cord. Demyelination and the resulting formation of this scar tissue in the CNS impair saltatory conduction along axons that is necessary for normal functioning of nerve impulses.

MS is a clinically heterogeneous disease that varies according to the location of plaques or lesions in the CNS. Recent pathological studies of lesions suggest that MS is an overlapping spectrum of related disorders [(1); (2); (3)]. Common symptoms include visual disturbances, loss of balance and coordination, spasticity, sensory disturbances, bladder and bowel incontinence, pain, weakness, fatigue, and paralysis. This debilitating disease also causes cognitive impairment in an estimated 45-65% of patients—with symptoms ranging from language deficits to bradyphrenia. Despite the substantial impairment and deterioration seen in MS, life span of affected individuals is only slightly

* Chapter adapted from:

Kenealy, S.J., Pericak-Vance, M.A., Haines, J.L. (2003) The genetic epidemiology of multiple sclerosis. *J Neuroimmunol* **143(1-2)**: 7-12.

shortened—creating a significant impact on quality of life for patients and on our nation's health care system.

The disease course of MS varies considerably among affected individuals. Cases may be episodic or progressive, severe or mild, and disseminated or primarily affecting the spinal cord and optic nerve. Although the disease has a broad range of age at onset (85% of cases occur between the ages of 14 and 55), initial symptoms typically present in early adulthood (between the ages of 20 and 40). MS occurs two to three times more frequently in women than men and is estimated to afflict approximately 400,000 people in the United States alone [(4)]. In most Caucasian populations, MS is second only to trauma as a cause of acquired neurologic disability arising in early to middle adulthood.

The diagnosis of MS is generally one of exclusion to eliminate conditions that mimic symptomology of the disease (e.g. B_{12} deficiency, AIDS, rheumatoid arthritis, systemic lupus erythematosus, Sjögrens syndrome, sarcoidosis, Lyme disease, adrenoleukodystrophy, and MELAS) [(5)]**.** In addition to clinical criteria, magnetic resonance imaging (MRI), evoked potential recordings, and cerebrospinal fluid examination can be used to confirm clinical diagnosis.

The course of MS is divided into two main subtypes: relapsing-remitting and primary progressive. The relapsing-remitting subtype is more common, characterized by two or more separate episodes of worsening symptoms involving different sites of the CNS, each lasting at least 24 hours and at least 1 month apart. Many relapsing-remitting cases

cease to remit and exhibit progression of at least one symptom in a slow or step-wise manner over at least 6 months. This course of the disease is referred to as secondary progressive MS. The second major subtype, primary progressive MS, is a less common form characterized by slow onset and steadily worsening symptoms involving sites of the CNS that do not remit from initial onset.

Though little is known about the underlying etiology, MS is physiologically an inflammatory disorder that results from an autoimmune response directed against CNS antigens—particularly myelin proteins. MS exhibits several characteristics common to autoimmune disorders—including evidence of environmental risk factors, increased frequency in women, partial susceptibility conferred by a human leukocyte antigen (HLA)-associated gene, and polygenic inheritance (the basis of the studies presented in this dissertation) [(6)]. Though little is known about the genetics of autoimmune disorders, the major histocompatibility complex (MHC), and more specifically the class II HLA genes, have been identified through candidate and/or genomic screen approaches as a genetic factor in several of these diseases. Class II MHC molecules, such as HLA, normally function to bind and present peptide antigens to antigen-specific T cells. It is thought that the dysregulation of this process in MS results in damage to the myelin sheath, producing the pathophysiological phenotype seen in the disease.

The clinical heterogeneity and complex etiology of MS have been confounding factors for studies of the disease. Yet despite these factors, it is clear that genes play a vital role

in disease susceptibility. The remainder of this chapter provides an overview of genetic studies for MS and a prelude to the research presented in this dissertation.

Genetic Epidemiology of MS

Epidemiological studies provide strong evidence for both environmental and genetic risk factors in MS. Numerous population and family-based studies have been conducted to assess disease prevalence and aggregation in an attempt to identify and elucidate genetic contribution to the disease.

Population Prevalence

Despite the disparity between prevalence rates cited in population-based studies for MS (values range from 0.88 to 224 per 100,000), there is a general consensus among researchers concerning a few observations [(7); (8)]. One observation is that the population prevalence of MS increases with distance from the equator. It is postulated that this distribution can be explained in part by both environmental factors (e.g. diet or vitamin D abnormalities) and population-specific genetics. Another observation is that the reported incidence of MS has increased over time. However, even these general conclusions should be cautiously interpreted due to the limited sample sizes in many studies, the changes in criteria and diagnosis that have accompanied improvements in health care, and the changes in epidemiological methods over time. Larger epidemiological studies will be required to definitively assess the prevalence and distribution of the disease.

Familial Aggregation

The involvement of genetic factors in MS has been demonstrated in numerous sibling risk, adoption, and twin studies. A commonly used measure in these studies is recurrence risk ratio (λ)—a value generated by comparing recurrence rates in the relatives of MS patients to the disease prevalence for the general population (9). Numerous familial aggregation studies have shown that the recurrence risk ratio for MS decreases with the degree of relationship between individuals. For example, studies have reported an increased relative risk (λ) of 100-190 in identical twins, 20-40 in full siblings, 7-13 in half siblings, and 5.5 in the offspring of an affected parent with MS [(10); (11); (12); (13)]. Compared to the general population, these elevated risks suggest a strong but non-Mendelian inheritance of MS susceptibility.

Twin studies from several populations indicate increased concordance rates among monozygotic (25-30%) compared to dizygotic (2-5%) twins with MS [(10); (11); (14)]. While these data also provide evidence of a strong genetic component to the disease, a monozygotic twin concordance rate significantly less than 100% also highlights the contribution of gene-environment interactions to MS disease susceptibility.

Evidence for genetically determined familial aggregation is also seen in adoption studies documenting an increased risk of MS only in biological relatives of adopted probands [(15)]. In addition, studies demonstrating similar risks for half siblings raised together and apart suggest the action of genetic rather than environmental factors in the disease [(12)].

Taken together, these epidemiological studies provide overwhelming evidence in support of a strong genetic component in MS. The data also suggest that, like most common complex diseases, MS susceptibility is the result of multiple genes acting either independently or interactively in their contribution to overall risk. The genetic etiology of MS may be a mixture of rare variants with strong environmental influences on risk and more common variants with modest influences on risk. Such heterogeneity would be similar to that seen in other complex neurodegenerative disorders, such as Alzheimer disease and Parkinson disease [(16); (17); (18); (19); (20)].

Approaches for Gene Identification

Functional Candidate Gene Studies

Functional candidate studies assess genes that are selected based on their potential biological relevance to a disease. Because MS is an autoimmune disease characterized by demyelination within the CNS, functional candidate genes such as those coding for immunoglobulin, cytokines, chemokines, T-cell receptors (TCR), interleukin, myelin antigens, and the human leukocyte antigen (HLA) have been investigated. However, with the exception of HLA, no functional candidates have consistently demonstrated association with MS.

Association between MS and class I HLA alleles was first reported in 1972 [(21); (22)]. Subsequent studies demonstrated that class II HLA alleles were more strongly associated with the HLA-DR2 haplotype [(23); (24)]. A majority of MS studies have focused on

Caucasian populations of northern European descent, where predisposition to MS is associated with the HLA-DR2 allele (more specifically, the HLA-DRB1*1501-DQA1*0102-DQB1*0602 haplotype). However, studies in additional populations have failed to replicate association with a particular allele or haplotype in the MHC (25). A recent study of the MHC was conducted in an African-American MS data set to capitalize on the haplotypic diversity and distinct LD patterns in the African-American population. A selective association was identified with HLA-DRB1*1501 in the study data set—suggesting a role for this locus independent of HLA-DQB1*0602 [(25)]. Several research groups continue to investigate the contribution of a gene or genes in the MHC to MS.

Aside from studies of the MHC, screening for functional candidate genes has been largely disappointing. Despite reports of numerous genes with significant results, most candidates have failed to be replicated in independent data sets. For example, an obvious candidate for MS, myelin basic protein (MBP), yielded both positive linkage and association results in a genetically isolated population in Finland [(26)]. However, other research groups have failed to replicate this result in non-Finnish populations [(27); (28); (29); (30); (31); (32); (33)]. Candidate genes identified through functional studies using the experimental autoimmune encephalomyelitis (EAE) mouse model for MS have also yielded inconsistent results.

Because the strength of the functional candidate gene approach lies in the knowledge of gene function, an improved understanding of autoimmunity and advances in the diagnosis

of MS will be necessary to improve this approach in the future. New methodologies, such as whole genome gene expression arrays and proteomics, will also add a considerable amount of information to aid in the selection of relevant functional candidates [(34); (35);(36); (37)].

Genomic Screens

Genomic screens test for genetic linkage of a trait to polymorphic markers spread throughout the genome. Numerous research groups have conducted genomic linkage screens for MS in an attempt to identify regions that harbor MS loci [(38); (39); (40); (41); (42); (43); (44); (45); (46); (47); (48); (49)]. However, the lack of replication of results from these studies has also been problematic. For example, while four initial genomic screens for MS identified over 70 regions of interest, little overlap is seen between these studies [(38); (39); (40); (41)].

The strongest and most consistent finding for linkage in MS studies is chromosome 6p21, the location of the MHC containing HLA. To date, the MHC is the only region that clearly and consistently demonstrates linkage and association with the disease. The MHC has been estimated to account for 10-50% of the genetic component of MS susceptibility, at least in Caucasians of northern European descent [(50); (24)]. It appears that the association with the HLA-DR2 allele explains this linkage signal, although this issue has been debated [(24); (51); (6)]. The exact mechanism by which a gene or genes in the MHC increase(s) disease risk has yet to be determined.

Despite the large number of genome-wide linkage studies that have been conducted, a significant proportion of the genetic contribution to MS is still unaccounted for. Several new approaches are being used to address the inadequate power of traditional linkage analysis to identify or verify MS loci.

One of the main difficulties in assessing genomic screen data for complex diseases is the lack of replication between studies. Genomic screens are intentionally designed to accept high false-positive rates in the interest of maintaining power to detect true loci. The replication of genomic screen results is therefore crucial for verification of genetic effects. As for many complex genetic diseases, replication of results has proved to be a formidable task in MS studies. For example, despite the fact that the MHC has shown the strongest genetic effect in MS thus far, not all linkage studies have replicated even this finding.

The lack of replication between linkage studies for MS suggests that the existence of genes with strong individual effects is unlikely. In addition, the use of different data sets (and data set structures), markers, and statistical approaches must be taken into account when comparing the results of both linkage and association studies for the disease.

Another confounding factor in linkage analysis of complex diseases is genetic heterogeneity—the presence of different mutations that produce similar disease phenotypes. Underlying genetic heterogeneity in MS likely masks the effects of true loci in many linkage studies. An approach that attempts to overcome loss of power due to

heterogeneity utilizes conditional analysis with covariates. By accounting for regions with known linkage (e.g. the MHC) using stratification or weighting schemes, conditional analysis can identify additional regions of interest. Conditional analysis can also provide evidence for interactive effects of loci, potentially providing increased power for detection of epistatic effects in complex diseases such as MS. The study design presented in this dissertation assesses a novel method for addressing genetic heterogeneity and utilizes several conditional analyses methods to investigate linkage follow-up regions.

A relatively new approach for positional mapping that minimizes some of the difficulties experienced in linkage studies is whole genome screening for linkage disequilibrium (LD). In an attempt to attain a higher degree of resolution than provided by traditional linkage analysis, Sawcer et al. conducted the first whole genome association study for MS. Using a DNA pooling scheme to screen 811 microsatellite markers at 0.5 cM intervals, this study reported significant association with ten markers, including those located at previously reported regions on chromosomes 1p, 6p (the location of HLA), 17q, and 19q [(52)]. Several additional groups have subsequently published whole genome screens for linkage disequilibrium using sample pooling schemes and case-control approaches [(53); (54); (55); (56); (57); (58); (59); (60); (61); (62); (63); (64); (65)].

The obvious disadvantage of whole genome studies for LD is the large number of polymorphic markers necessary for sufficient coverage of the genome. Collaborating

scientists from several countries are currently participating in an International HapMap Project to document common patterns of variation in the human genome. This project aims to identify single nucleotide polymorphisms (SNPs), haplotype blocks, and haplotype tagging SNPS (htSNPs) in order to provide scientists with resources to more efficiently investigate variation throughout the genome. In addition, several new genotyping methods and platforms have been developed to allow for high-throughput data generation necessary to conduct these studies. Companies such as Parallele Biosciences (the MegaAllele™ system), Illumina (the Sentrix® BeadChip system), and Affymetrix (the GeneChip® system) currently provide high-throughput genotyping products and services for whole genome studies (see Parallele Biosciences, Illumina, and Affymetrix websites). The continued development of cost-effective genotyping methods will make whole genome approaches likely candidates for future investigation of genes with moderate effects on MS susceptibility.

In conjunction with the International Multiple Sclerosis Genetics Consortium, we recently published a high-density linkage screen for MS using the Illumina BeadArray linkage mapping panel [(66)]. Preliminary multipoint linkage analyses of 730 multiplex families of northern European descent revealed strongest evidence of linkage to chromosomes 6p21 (the location of HLA), 5q33, and 17q23. Ordered subset analyses provided additional evidence of linkage to a locus on chromosome 19q13 that acts independently of the MHC. Additional analyses are ongoing to identify homogenous subsets and investigate gene-gene interactions.

Locational Candidate Gene Studies

Traditional locational candidate genes are selected from chromosomal regions identified through genomic screens or chromosomal abnormalities in affected patients (e.g. duplications, deletions, or translocations). Although locational candidate methods initially involved a random screening process, advances in genomic mapping have allowed for modified genetic approaches that incorporate positional and candidate gene methodologies. For example, following identification through genomic screens, regions of interest can be scanned for functional candidate genes using information from several public and private databases. In conjunction with family-based association methods, this approach allows for a more directed investigation of genes.

One example of this approach involves the chromosome 19q13 region. Despite the fact that this region has been identified in several genomic screens, the gene responsible for linkage on 19q13 remains to be definitively identified. One of the candidate genes in this region identified through bioinformatics is apolipoprotein E (APOE). The APOE gene codes for a major lipid carrier protein (apoE) in the brain. The apoE protein has long been associated with regeneration of axons and myelin following the formation of lesions in the central and peripheral nervous tissue. Decreased apoE concentrations in cerebrospinal fluid (CSF) in MS patients compared to healthy controls have been reported, and a corresponding decrease in intrathecal apoE synthesis may influence the degree of MS exacerbation over time. Although studies of APOE have consistently shown no effect on MS risk, association of the APOE-4 allele with increased disease progression or severity of disease course have been reported and confirmed in multiple

studies [(67); (68); (69); (70); (71); (72); (73); (74)]. Investigation of the nearby poliovirus receptor related protein 2 (PVRL2) has also revealed association with MS disease course, while apolipoprotein C2 (APOC2) and immunoglobulin-like transcript 6 (ILT6) have demonstrated association with MS susceptibility [(75); (76); (77); (78)]. Examination of a large well-phenotyped MS data set will be necessary to construct more detailed LD maps of the region and identify the gene(s) and polymorphism(s) responsible for the genetic effect on chromosome 19q13.

Future Directions in MS Research

While it is clear that MS is a disease of oligogenic etiology, identifying specific genes has been difficult. With the exception of HLA, linkage analysis and candidate gene approaches have demonstrated insufficient power to identify other genes or epigenetic factors that modulate MS disease expression. New approaches and methodologies will be necessary to identify the remaining genetic effect in this complex disease.

At the population level, several methods have been used to test candidate genes for an effect in MS. Initial studies primarily used an approach that compared allele frequencies for polymorphisms in case versus control groups ("case-control studies"). However, case-control studies are sensitive to sample size, stringency of diagnosis, and appropriate matching of controls. Population admixture in improperly matched controls can lead to spurious association results that are indistinguishable from results arising from true genetic effects.

In an attempt to overcome these confounding factors, family-based association methods have been developed. These approaches require only one affected individual and their parents (a "trio") or one affected individual and at least one of their unaffected siblings (a "discordant sib-pair" or "DSP") to serve as well-matched controls. The most common method of family-based association is the transmission disequilibrium test (TDT) [(79)]. Variations such as the sibling transmission disequilibrium test (sib-TDT) and pedigree disequilibrium test (PDT) have also been developed to allow for sampling of a variety of control groups [(80); (81)]. In addition, the genotype pedigree disequilibrium test (geno-PDT) was developed to test for association with particular genotypes [(82)]. Because our data set consists of a variety of a variety of family structures, the PDT was used to assess disease-marker disequilibrium. The PDT is an extension of the TDT that allows for analysis of data from related nuclear families and discordant sibships within extended pedigrees. The standard for significance in our PDT analyses was a p value ≤ 0.05.

By tracking the transmission of gametes, the TDT and its variants detect association only between linked loci—providing tests of both linkage and association. These tests use nontransmitted alleles as controls, eliminating spurious association results caused by population substructure. Because the power of family-based association methods arises from reliance on relatively small intervals of linkage disequilibrium (LD), this approach serves as a complimentary method to broader linkage analyses methods using genomic screens.

These association studies seeking to identify loci of moderate effect in MS will require a set of markers present in greater density and with greater ability to detect LD than microsatellites. Because of their frequency, stability, and amenability to automation for high-throughput analysis, SNPs are rapidly becoming the standard marker for such association studies. The large number of assays and new high-throughput methods for data generation using SNPs allow for the construction of detailed haplotype blocks. The knowledge of regional LD and marker information from several databases can also aid in efficient selection of appropriate markers for association studies.

To attain sufficient power for detection of MS loci using current linkage and association methods, we have pooled large multiplex, simplex, and case-control data sets with our collaborators in the Multiple Sclerosis Genetics Group (MSGG). These data sets are a crucial resource for performing conditional analyses to identify phenotypic and genotypic subsets of the disease. These data sets may also allow for independent replication of interesting preliminary findings.

Despite the overwhelming evidence for genetic involvement in MS, much of the genetic effect remains to be identified or elucidated. Recent advances in genotyping and statistical analysis methods are providing researchers with the tools necessary to address the challenges involved in identifying genes for complex genetic diseases. The following dissertation presents a genomic convergence approach that incorporates a variety of new methods and statistical tools to conduct a directed investigation of genetic contribution to MS. This genomic convergence approach led to the investigation of several

chromosomal regions, with the most promising evidence of a genetic locus conferring susceptibility to MS on chromosome 1q44.

CHAPTER II

HYPOTHESIS AND SPECIFIC AIMS

General Hypothesis: There are genes underlying the susceptibility to multiple sclerosis.

Specific Aims:

1. **Conduct a simulation study using families generated with Genometric Analysis Simulation Program (GASP) software to assess the effectiveness of using haplotype-based positional mapping to define a minimum candidate region for a disease of interest.** Several variables (e.g. sample size, pattern of inheritance, and heterogeneity) will be investigated for their effect on the power of this approach. (REFER TO CHAPTER III)
2. **Test candidate genes for association with MS:**
    a. **Identify and select a genetic interval of interest for MS.** Potential intervals of interest will identified by comparing positive results generated in genomic screens conducted for MS and other autoimmune disorders. (REFER TO CHAPTER IV)
    b. **Apply the approach from Specific Aim 1 to the interval of interest identified in Specific Aim 2a.** Follow-up with microsatellite markers and single nucleotide polymorphisms (SNPs) will be performed prior to recombination breakpoint analysis in the selected region of interest. (REFER TO CHAPTERS V AND VI)

c. **Identify candidate genes in the selected interval of interest.** Candidate genes will be identified based on potential biological relevance to MS and/or involvement in common physiological pathways of autoimmune disorders. (REFER TO CHAPTER VII)

d. **Select a region and a narrow interval of interest for follow-up with SNPs in Specific Aim 3.** (REFER TO CHAPTER VII)

3. **Measure association between a dense population of SNPs and MS in the region of interest identified in Specific Aim 2.**

a. **Prioritize SNP markers for an MS association study in the interval of interest based on conservation between human, mouse, rat, and chick genome sequences.** The WebMCS tool will be used to identify multi-species conserved sequences in the chromosomal region of interest by integrating comparative information from the orthologous mouse, rat, and chick genomic sequences. Conserved regions will be scanned for SNPs in the public databases and selection of SNPs for genotyping in Specific Aim 3b will be based on several criteria: informativeness, validation, location, putative function, and Illumina assay score. (REFER TO CHAPTER VI)

b. **Measure association between a dense population of SNPs located in conserved regions of interest and a data set of families linked to the region.** High-throughput genotyping will be performed on the Illumina Bead Array platform for 768 SNPs selected from conserved regions identified in Specific Aim 3b. A data set of ~200 multiplex families will be tested for association with these SNPs using the Pedigree

Disequilibrium Test and Haploview. [SNPs demonstrating the strongest evidence for association in the multiplex families will be further tested in at least one of several available simplex family data sets.] (REFER TO CHAPTER VI)

CHAPTER III

HAPLOTYPE-BASED POSITIONAL MAPPING[†]

Introduction

Numerous genomic screens have been conducted in an attempt to identify putative genes for both Mendelian diseases and complex genetic disorders. Because the typical marker interval in a genomic screen is ~ 10 cM, subsequent studies are often required to narrow chromosomal regions of interest to a reasonable size for candidate gene or fine mapping association studies. In monogenic disorders, haplotype analysis methods have been widely used to identify minimum candidate gene regions. Techniques for narrowing linkage signals in these studies are based on the expectation that affected individuals will consistently inherit a relatively small region containing the disease locus. However, extensive heterogeneity, gene-gene interactions, and small family size in typical data sets complicate this simple assumption for many complex diseases. These confounding factors make identification of consistently inherited regions in complex diseases unlikely—resulting in failure to narrow linkage intervals to a practical size for subsequent association studies.

The focus of the study presented in this chapter is the investigation of a methodology formulated to address locus heterogeneity. Locus heterogeneity is the presence of two or

[†] The GASP simulation project was performed in collaboration with Tricia A. Thornton-Wells.

more loci that lead to the same clinical phenotype in different families. The implication of this heterogeneity in linkage analysis is the reduction of power to detect true signals originating from only a subset of families. One approach that has been used to address the statistical difficulties posed by locus heterogeneity is *a priori* subsetting of families using phenotype data. However, this approach is based on the assumption that multiple loci produce different and distinguishable phenotypes and is therefore unlikely to work for the complex phenotype and clinical diagnosis of MS. Better methods for narrowing linkage intervals in the presence of locus heterogeneity are clearly needed for studies of MS and other complex genetic diseases.

An approach that has been used to narrow linkage intervals for Mendelian diseases is consensus haplotyping. In this approach, genomic screen data is used to reconstruct familial haplotypes. Recombination breakpoint analysis is then performed in all families to identify a consensus region(s) that will be further investigated by genotyping additional locational or functional candidate markers. Although this approach has been successful in identifying minimum candidate regions for monogenic diseases, the approach was not utilized for complex disorders until recently [(83)].

In this recent study, Hutcheson et al. modified the consensus haplotype approach for application to an Autism data set. In order to use the recombination breakpoint method in this data set, the authors relaxed the requirement of consistent inheritance of haplotypes across all families. They postulated that for sib-pair data sets, 1/4 of families demonstrate sharing between siblings for any given marker by chance alone. Standard linkage

analysis methods cannot distinguish these families linked by chance from families that exhibit true linkage. To estimate the expected proportion of families that are truly linked to a given region, the authors performed the calculation shown in Table 1. In the example illustrated in this table, it is assumed that 1/3 of families from a given data set carry a risk allele in a particular chromosomal region. Of the 2/3 of families that do not carry this risk allele, 1/4 (or 1/6 of the overall data set) will demonstrate linkage to this region by chance alone. Families demonstrating linkage will therefore account for 1/3 + 1/6, or 1/2, of the overall data set. However, of these apparently linked families, only 2/3 of families actually carry the risk allele.

Table 1. Calculation for the Expected Proportion of Linked Families [Adapted from Hutcheson et al. (83)]

	Susceptibility Allele Present	**Susceptibility Allele Absent**
Prior Probability	1/3	2/3
Conditional Probability	1	1/4
Joint Probability	1/3	1/6
Posterior Probability	$\left(\frac{1/3}{(1/3+1/6)}\right) = 2/3$	$\left(\frac{1/6}{(1/3+1/6)}\right) = 1/3$

The results of this calculation suggest that a majority of apparently linked families provide consistent localization of the risk allele, and that the proportion of truly linked families can be calculated and used to define consistently inherited segments. For the

example provided in Table 1, boundaries for the consistently inherited segment would be selected where at least 2/3 of families demonstrate linkage.

With this modified method, the authors were able to perform recombination breakpoint analysis on maternally and paternally derived chromosomes in their Autism data set. The result of this approach was narrowing of a critical region on chromosome 7 from 34 cM to 6 cM. Families linked to the 6 cM region were then used to perform additional genotyping and recombination breakpoint analysis—resulting in the identification of a 3 cM interval that was considerably more amenable for subsequent association studies.

Specific Aim 1 addresses the power of this modified approach to narrow a minimum candidate region for a complex disease. The Genometric Analysis Simulation Program (GASP) was used to simulate data sets to assess power and the effect of specified variables on this approach. Studies of simulated data were followed by a "proof-of-principle" analysis with genotype information from a region demonstrating linkage in our MS data set.

Material and Methods

GASP software was used to generate simulated data sets with several specified parameters (see GASP website). One hundred data sets were generated per disease model for nuclear families consisting of two unaffected parents and three offspring (with at least two of the offspring being affected) (Figure 1).

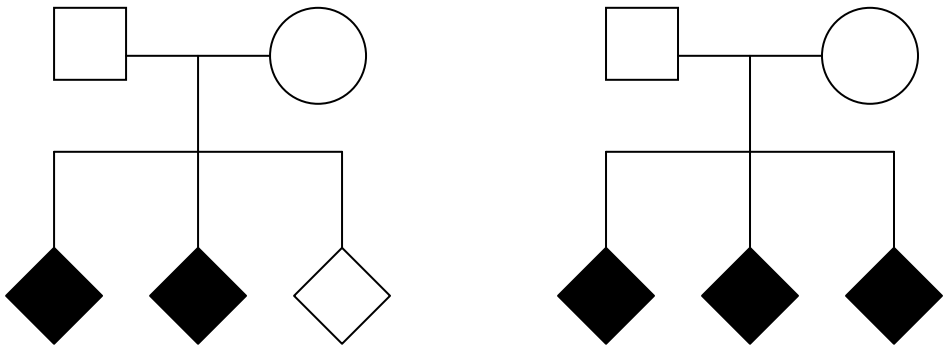


Figure 1. GASP Pedigree Examples

Ten biallelic markers with minor allele frequencies of 0.50 were simulated at 10 cM intervals along a chromosome to mimic the study design of a traditional genomic screen. A disease locus was simulated halfway between two of the markers. A second unlinked disease locus was also simulated to mimic genetic heterogeneity (Figure 2).

Data sets were generated to demonstrate 50% locus heterogeneity (50% locus A; 50% locus B) and 25% locus heterogeneity (25 locus A; 75% locus B). Linkage analyses were performed under a recessive disease model with a disease prevalence of 0.10. Two-point LOD scores were calculated in FASTLINK and heterogeneity LOD (HLOD) scores were calculated in HOMOG [(84); (85); (86); (87)].

Following linkage analysis of data sets for each disease model, the marker generating the highest LOD score was identified. Families demonstrating linkage to any marker within 20 cM of this peak LOD score marker were selected for recombination breakpoint

analysis. Haplotypes were constructed for linked families using SIMWALK version 2.9 [(88)]. Each family was investigated to identify which loci demonstrated sharing on both haplotypes in all affected offspring (Table 2). Blocks of loci demonstrating sharing in at least 2/3 of families were used to define the critical region in each data set.

The power of this method was determined by calculating the percentage of data sets that included the true disease locus within the identified critical region. Significant power was reached with ≥ 80% of data sets achieving this standard. The correlation between power and size of the critical region was also investigated.

The same recombination breakpoint method, linkage analyses, and power calculations were also performed on a set of follow-up markers spaced at 2 cM intervals in the selected region to mimic the design of a traditional follow-up study (Figure 2).

Following assessment of the recombination breakpoint method in simulated data sets, the method was also applied to an existing data set for MS in the 1q44 region. MS genotyping data was generated as described in Chapter 5. Families demonstrating linkage to any marker in the follow-up region (i.e. within 20 cM of peak LOD score marker D1S1634) and containing genotyping data for two unaffected parents and at least two affected offspring were selected for recombination breakpoint analyses. In the data set of 91 families linked to the 1q44 region, 57 families demonstrated consistent sharing on the paternal alleles in all affected individuals, while 59 families demonstrated consistent sharing on the maternal alleles in all affected individuals.

Linked Chromosome:

Stage 1 Markers: **Disease Locus A**

1 2 3 4 5 * 6 7 8 9 10

Stage 2 Markers:

1 2 3 4 5 * 6 7 8 9 10

cM Location:

-45 -35 -25 -15 -5 5 15 25 35 45

Unlinked Chromosome:

Disease Locus B

*

Figure 2. Simulated Markers Chromosome images were obtained from the Genome Database (see GDB website).

Table 2. Example Data Set Loci demonstrating sharing on both haplotypes in all affected offspring are denoted by "x".

Marker #	Simulated Family #															% Sharing
	484	746	2225	2534	4218	5311	5841	6225	6250	6335	7492	8955	9785	10046	10610	
1 (@ -45 cM)		x				x	x		x			x	x	x		47 %
2 (@ -35 cM)		x				x	x	x	x			x	x	x		53 %
3 (@ -25 cM)		x			x	x	x	x	x			x	x	x		60 %
4 (@ -15 cM)		x			x	x	x	x	x			x	x	x		60 %
5 (@ -5 cM)	x		x		x	x	x	x	x		x	x	x	x		73 %
6 (@ +5 cM)			x	x	x		x	x	x	x	x	x	x	x	x	80 %
7 (@ +15 cM)			x	x	x		x	x	x	x	x	x	x	x	x	80 %
8 (@ +25 cM)			x	x	x		x		x	x	x		x	x	x	67 %
9 (@ +35 cM)			x		x		x		x		x		x	x	x	53 %
10 (@ +45 cM)			x		x		x	x	x				x	x	x	53 %

The proportion of families demonstrating sharing among all affected offspring was calculated for each marker in the 1q44 region. The utility of the recombination breakpoint method was determined by assessing the ability of this method to narrow the linkage interval on chromosome 1q44 compared to other available methods (e.g. subsetting, conditional analysis, and ordered subset analysis). The recombination breakpoint method was also performed using microsatellite genotypes to assess the effect of more informative markers on the ability to detect sharing in the MS data set (D1S1594, D1S547, and D1S1634).

Results

Plots containing HLOD scores for the 10 cM interval and 2 cM interval simulated data sets are provided in Figures 3 and 4, respectively. In the simulated data, the recombination breakpoint method had ~ 75% power to correctly localize the disease locus within a 40 cM region and ~ 60% power within a 30 cM region using the 10 cM marker interval design. In contrast, the recombination breakpoint method had only ~ 20% power to correctly localize a disease locus within a 10 cM region using the 2 cM marker interval design. The results of power calculations for identification of the disease locus in each disease model and study design are provided in Tables 3 and 4.

Proportions of families demonstrating paternal, maternal, and combined haplotype sharing in the MS data set are provided in Figures 5-7, respectively.

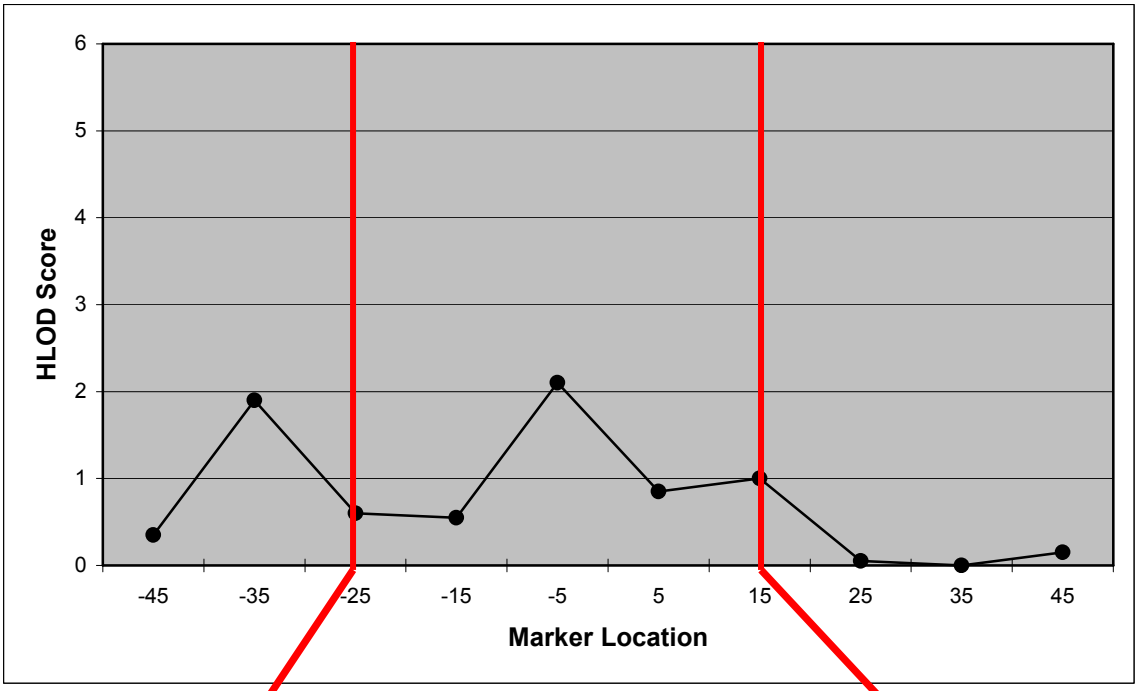


Figure 3. HLOD Scores for 10 cM Spaced Markers

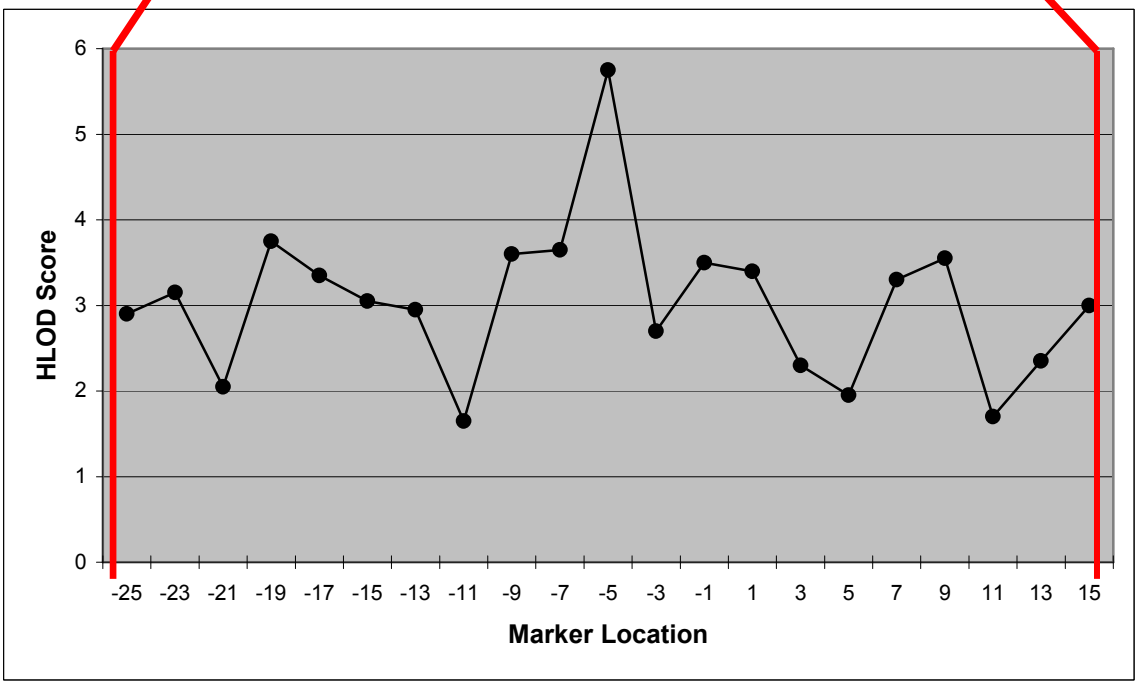


Figure 4. HLOD Scores for 2 cM Spaced Markers

Table 3. Power to Identify the Disease Locus (Locus A) in the 10 cM Map

Heterogeneity	Range		
	20 cM	**30 cM**	**40 cM**
50% Locus A; 50% Locus B	> 20%	> 50%	> 75%
25% Locus A; 75% Locus B	> 30%	> 55%	> 75%

Table 4. Power to Identify the Disease Locus (Locus A) in the 2 cM Map

Heterogeneity	Range		
	10 cM	**16 cM**	**20 cM**
50% Locus A; 50% Locus B	> 10%	> 35%	> 50%
25% Locus A; 75% Locus B	> 20%	> 40%	> 60%

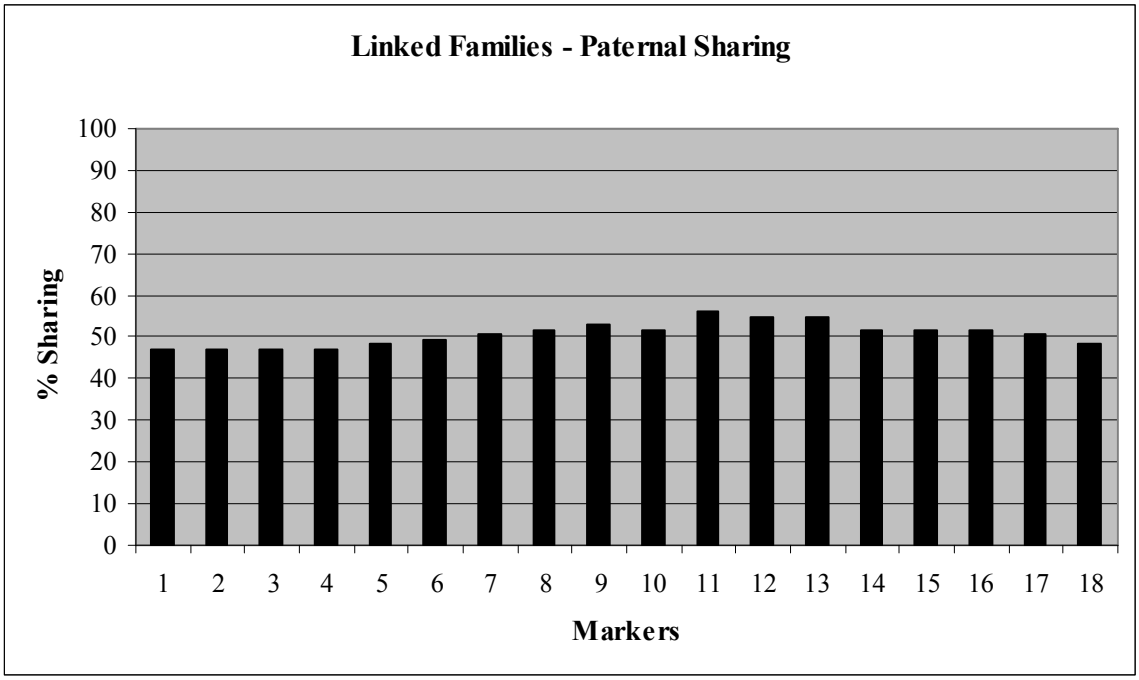


Figure 5. Proportion of Families Demonstrating Sharing in the MS Data Set (Paternal Haplotype)

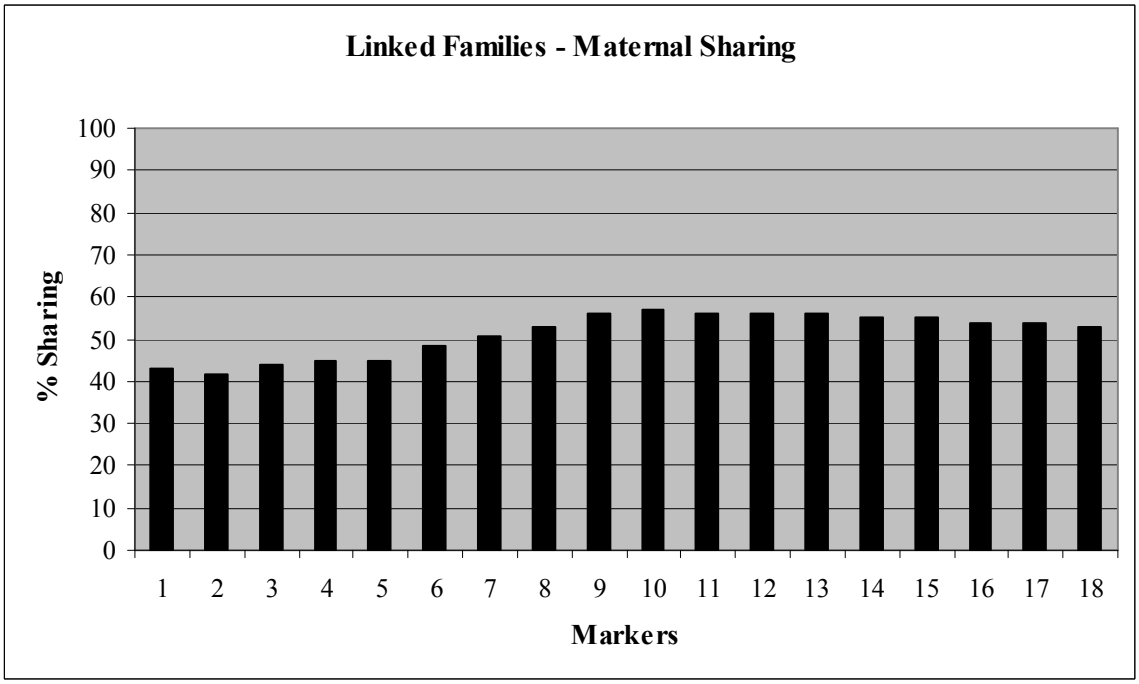


Figure 6. Proportion of Families Demonstrating Sharing in the MS Data Set (Maternal Haplotype)

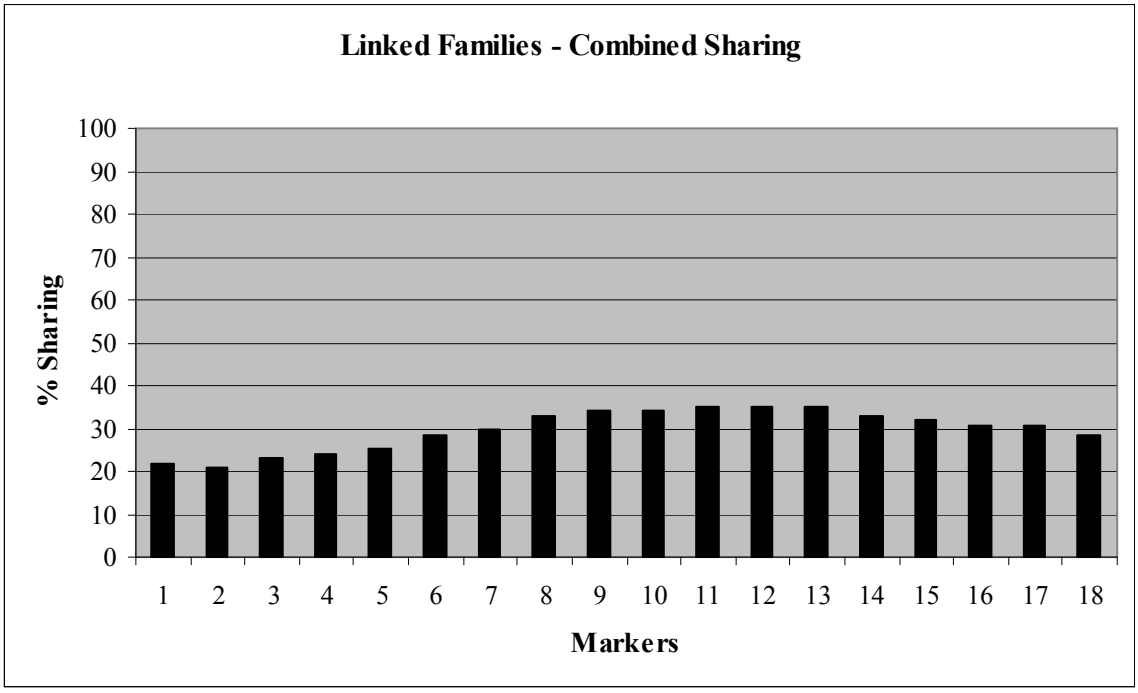


Figure 7. Proportion of Families Demonstrating Sharing in the MS Data Set (on both Paternal and Maternal Haplotypes)

Discussion

The modified consensus haplotyping approach demonstrated only modest power to narrow a minimum candidate region in the simulation data. Perhaps not surprisingly, the approach also demonstrated only a modest ability to narrow the minimum candidate region on chromosome 1q44 in MS genotyping data. Given the encouraging results generated in the Autism study by Hutcheson et al., these results were rather disappointing. While it is not possible to definitively determine the source of variation between studies, one difference between the Hutcheson et al. study design and our study design was the use of microsatellite markers rather than SNPs. To assess whether marker type affected the results in our MS data set, we performed the consensus haplotyping

approach with genotype data from the original SNPs and three microsatellite markers spaced at ~ 10 cM intervals in the 1q44 region. The inclusion of microsatellite genotypes only slightly increased evidence of sharing and failed to further narrow the minimum candidate region.

Fortunately, more encouraging results for narrowing minimum candidate regions in the MS data set were simultaneously being generated in preliminary analyses of genotyping data for Specific Aim 2. For example, in contrast to the modified haplotype approach, ordered subset analysis (OSA) of the 1q44 region in Specific Aim 2b successfully narrowed the critical linkage interval to ~ 3.5 Mb for a LOD score cut-off of 3.5 (corresponding to a – 1.0 LOD score confidence interval) and ~ 7.0 Mb for a LOD score cut-off of 2.5 (corresponding to a – 2.0 LOD score confidence interval). Covariate analyses, including OSA, were therefore selected as the method to address genetic heterogeneity in studies of the MS data set in Specific Aim 2 (see Chapter 5).

Although the modified haplotype approach failed to demonstrate substantial power to identify a minimum candidate region in our studies, this approach should not yet be abandoned as a potential method for other studies. Additional disease models and data sets should be assessed to further investigate the utility of this method in addressing genetic heterogeneity in complex genetic diseases.

CHAPTER IV

A SECOND-GENERATION GENOMIC SCREEN FOR MULTIPLE SCLEROSIS[‡]

Abstract

Multiple sclerosis (MS) is a debilitating neuroimmunological and neurodegenerative disorder. Despite substantial evidence for polygenic inheritance, the MHC is the only region that clearly and consistently demonstrates linkage and association in MS studies. The goal of this portion of the study was to identify additional chromosomal regions that harbor susceptibility genes for MS. With a panel of 390 microsatellite markers genotyped in 245 U.S. and French multiplex families (456 affected relative pairs), this is the largest genomic screen for MS conducted to date. Four regions met both of our primary criteria for further interest (HLOD and Z scores > 2.0): 1q (HLOD = 2.17; Z = 3.38), 6p (HLOD = 4.21; Z = 2.26), 9q (HLOD = 3.55; Z = 2.71), and 16p (HLOD = 2.64; Z = 2.05). Two additional regions met only the Z score criterion: 3q (Z = 2.39) and 5q (Z = 2.17). Further examination of the data by country (U.S. and France) identified one additional region demonstrating suggestive linkage in the U.S. subset (18p: HLOD = 2.39) and two additional regions generating suggestive linkage in the French subset (1p: HLOD = 2.08; and 22q: HLOD = 2.06). Examination of the data by HLA-DR2

[‡] Chapter adapted from:

Kenealy, S.J., Babron, M.C., Bradford, Y., Schnetz-Boutaud, N., Haines, J..L, Rimmler, J.B., Schmidt, S., Pericak-Vance, M.A., Barcellos, L.F., Lincoln, R.R., Oksenberg, J.R., Hauser, S.L., Clanet, M., Brassat, D., Edan, G., Yaouanq, J., Semana, G., Cournu-Rebeix, I., Lyon-Caen, O., Fontaine, B. (The American-French Multiple Sclerosis Genetics Group) (2004) A second-generation genomic screen for multiple sclerosis. *Am J Hum Genet* **75(6)**: 1070-1078.

stratification identified four additional regions demonstrating suggestive linkage: 2q (HLOD = 3.09 in the U.S. DR2- families), 6q (HLOD = 3.10 in the French DR2- families), 13q (HLOD = 2.32 in all DR2+ families and HLOD = 2.17 in the U.S. DR2+ families), and 16q (HLOD = 2.32 in all DR2+ families and HLOD = 2.13 in the U.S. DR2+ families). These data suggest several regions that warrant further investigation in the search for MS susceptibility genes.

Material and Methods

Families

The data set used in this study consisted of families from a previous genomic screen conducted by the MSGG [(39)], 66 subsequently ascertained U.S. families, and 94 French families. The full data set included 245 multiplex families consisting of 587 affected individuals, 344 affected sib-pairs, 112 other affected relative pairs, and a total of 1085 samples (Table 1).

Table 1. Description of the Data Set Families were designated HLA-DR2+ if every affected individual carried at least one HLA-DR2 allele or HLA-DR2- if no affected individuals carried an HLA-DR2 allele.

	# Families	# Affecteds	# ASPs	# Other ARPs	# HLA-DR2+ Families	# HLA-DR2- Families
U.S.	151	383	242	88	83	31
French	94	204	102	24	28	35
All	245	587	344	112	111	66

U.S. families were ascertained by the University of California at San Francisco (UCSF). All U.S. affected family members were examined or had their medical records reviewed by a collaborating physician. Families were extended through all affected first-degree relatives if possible. French families were collected through a national network of university and community hospitals and private practitioners. All French affected family members were examined by a clinician from one of three centers (Paris, Rennes, or Toulouse).

All protocols were approved by the appropriate Institutional Review Boards and all individuals provided informed consent before participating in the study. Positive family histories were investigated by direct contact with other family members, request for medical records, and by clinical examination, laboratory testing, or paraclincial studies (MRI scanning and evoked-response testing). Consistent and stringent clinical criteria were applied as previously described [(5); (39)]. Individuals were placed into one of four categories: definite MS, probable MS, possible MS, and no evidence of MS. Only definite MS individuals were classified as affected individuals in the analyses.

To account for possible heterogeneity, the data were examined for differences by country (U.S. vs. France) and HLA-DR2 genotype (HLA-DR2+ vs. HLA-DR2). Families were designated DR2+ if every affected individual carried at least one HLA-DR2 allele or DR2- if no affected individuals carried an HLA-DR2 allele (Table 1).

Molecular Analysis

After obtaining informed consent, blood samples were collected from each study participant. Genomic DNA was extracted from blood samples using standard procedures as described elsewhere [(89)]. All DNA samples were coded and stored at 4°C prior to use.

Marker primer sequences were obtained from the Genome Database (see GDB website) or designed with Primer3 software (see Primer3 website) and synthesized by Invitrogen Life Technologies (Carlsbad, CA). Amplification was performed in a PCR Express machine (ThermoHybaid, Needham Heights, MA) with the following conditions: 94°C-4 min.; 94°C-15 sec., AT-30sec., 72°C-45 sec. (35 cycles); 72°C-4 min. PCR products were denatured for 3 min. at 95°C and run on a 6% polyacrylamide gel (Sequagel-6®from National Diagnostics, Atlanta, GA) for ~ 1 hr. at 75 W. Gels were stained with a SybrGold® rinse (Molecular Probes, Eugene, OR) and scanned with the Hitachi Biosystems FMBIOII laser scanner (Brisbane, CA). Genotypes for HLA-DR in the U.S. families were determined at UCSF using non-radioactive PCR-SSOP (Dynal, Norway). Genotypes for HLA-DR in the French families were determined using reverse dot blot hybridization.

Marker order and intermarker distance were determined using linkage reference maps (see Marshfield website; deCODE website). The average intermarker distance for the screen was < 10 cM. The Vanderbilt and Duke laboratories each genotyped a subset of markers on the complete set of DNA samples. Laboratory personnel were blinded to

pedigree structure, affection status, and location of quality control samples. Duplicate quality control samples (3 unblinded CEPH individuals and 4 blinded controls) were placed both within and across plates and equivalent genotypes were required to ensure accurate genotyping.

Allele frequencies were calculated from the genotyped founders in each family. Hardy-Weinberg calculations were performed for each marker and Mendelian inconsistencies were identified using PedCheck [(90)]. Suspect genotypes were re-read and/or re-run. All microsatellites were required to have > 85% of possible genotypes. Verification of relationships between pairs of samples within families was performed using RELPAIR [(91)]. Markers and samples failing to pass quality control measures were dropped from the analyses.

Statistical Analysis

Both model-based and model-free analyses were performed. Parametric (model-based) analyses were conducted using autosomal dominant and autosomal recessive models with disease allele frequencies of 0.01 and 0.20 (respectively) to model a common susceptibility allele. A penetrance value of 0.95 was used for both dominant and recessive models and individuals with no evidence of MS were coded as normal for these analyses. Two-point LOD scores were calculated in FASTLINK and heterogeneity LOD (HLOD) scores were calculated in HOMOG [(84); (85); (86); (87)]. Two-point HLOD scores for the overall data sets, HLA-DR2+ subsets, and HLA-DR2- subsets are provided in Figures 1, 2, and 3 (respectively).

Multipoint model-free analyses were performed using the "score pairs" option and the exponential model in Allegro [(92); (93)]. Multipoint results are given in terms of Z scores. Because the HLA-DR2 allele is known to be associated with MS susceptibility, potential interactions between HLA and other regions were tested by calculating correlation between pairwise family NPL values in the 236 nuclear families with at least one affected sib-pair.

The criterion to consider a chromosomal region as interesting was at least one marker with a maximum heterogeneity LOD (HLOD) score > 2.0 or a multipoint Z score > 2.0. Because other research groups have advocated using a more liberal criterion to identify regions of interest from genomic screens, we also report markers generating HLOD and/or Z scores > 1.5 and have made the complete set of HLOD and Z scores available at the Vanderbilt Center for Human Genetics website (see CHGR Supplemental Data website).

A number of statistical tests were performed on the microsatellite markers, disease models, and subsets—raising concern about multiple comparisons. The level of correction necessary to account for these factors is a topic of substantial debate and selecting an appropriate level of corrections is not clear. We therefore have chosen to present the results of this study without correction for multiple tests. To gain some idea of a significance level for our data set, we performed a simulation using the observed family structures. The value of the threshold for HLOD scores for a genome-wide type I

error of 1% was 1.86 under the hypothesis of no linkage. The value of the threshold for the model-free statistic Z for a genome-wide type I error of 5% was 3.56 under the hypothesis of no linkage.

Results

Overall Analysis

Four regions met both primary criteria for further interest (HLOD and/or Z scores > 2.0): 1q (HLOD = 2.17; Z = 3.38), 6p (HLOD = 4.21; Z = 2.26), 9q (HLOD = 3.55; Z = 2.71), and 16p (HLOD = 2.64; Z = 2.05) (Tables 2 and 3). Seven regions (1p, 2q, 6q, 13q, 16q, 18p, and 22q) generated only HLOD scores > 2.0 and two regions (3q and 5q) generated only Z scores > 2.0. Using a more liberal criterion of HLOD scores > 1.5 identified eight additional regions in the two-point analysis (2p, 3p, 3q, 4p, 4q, 7p, 12q, and 15q), while a more liberal criterion of Z scores > 1.5 identified three additional regions in the multipoint analysis (2p, 10q, and 18p).

Site Stratification

Further examination of the data by country (U.S. and France) identified three regions demonstrating suggestive linkage in the U.S. subset: 6p (HLOD = 3.30), 9q (HLOD = 2.32), and 18p (HLOD = 2.39) (Table 2). Two of these three regions were also identified in examination of the overall data set as discussed above: 6p and 9q. Site stratification also identified three regions generating suggestive linkage in the French subset alone: 1p (HLOD = 2.08), 16p (HLOD = 2.64), and 22q (HLOD = 2.06).

HLA-DR2 Stratification

Further examination of the data identified markers generating suggestive HLOD scores after HLA-DR2 stratification in seven regions: 2q (HLOD = 3.09 in the U.S. DR2- families), 6p HLOD = 2.24 in all DR2+ families), 6q (HLOD = 3.10 in the French DR2- families), 9q (HLOD = 2.05 in all DR2+ families), 13q (HLOD = 2.32 in all DR2+ families and HLOD = 2.17 in the U.S. DR2+ families), 16q (HLOD = 2.32 in all DR2+ families and HLOD = 2.13 in the U.S. DR2+ families), and 18p (HLOD = 2.25 in all DR2- families and HLOD = 2.84 in the U.S. DR2- families) (Table 2). Four of these regions (2q, 6q, 13q, and 16q) were identified only in HLA-DR2 stratified subsets.

Discussion

Genetic linkage analysis has proven to be successful in locating Mendelian disease genes, but whole genome screens have been less successful in locating genes for complex genetic diseases such as MS. Rarely does any region reach a single-stage significance level—indicating that a two-stage design requiring confirmation in at least one additional data set is necessary to declare linkage. Whole genome screens for MS have identified over 70 regions potentially harboring MS loci. However, with the singular exception of the MHC, there has been a lack of consensus across studies. Our results hold true with this general pattern, but several of the regions identified in this screen do recapitulate significant linkage suspected by other groups. Our large data set and stringent criterion for identifying regions of interest (HLOD and/or Z scores > 2.0) suggest several regions of linkage for MS.

Not surprisingly, a marker in the HLA-DR region on chromosome 6p21 generated the highest two-point LOD score (HLOD = 4.21) and one of the highest multipoint LOD scores (Z = 2.26) for the entire screen, confirming evidence of a risk factor in this region [(39); (40); (41); (46)].

The strongest evidence of linkage to a non-MHC region in the overall data set was for chromosome 9q34 (HLOD = 3.55; Z = 2.71). The initial Multiple Sclerosis Genetics Group (MSGG) screen, MSGG follow-up study, and recent screens in Nordic sib-pairs and Turkish families also demonstrate moderate support for 9q, highlighting the need for further investigation of this region [(39); (48); (46); (49)].

Another region of interest from the screen that is supported by several lines of evidence is chromosome 1q. Marker D1S547 in the 1q44 region met the criterion for further interest in both two-point and multipoint calculations (HLOD = 2.17; Z = 3.38). Nearby markers have also demonstrated suggestive linkage and/or association in several other MS screens that have been conducted in a variety of study populations [(44); (47); (52); (53); (94)]. Another compelling piece of evidence for chromosome 1q is linkage to this region in the autoimmune disorders rheumatoid arthritis and systemic lupus erythematosus, suggesting the presence of a gene for general autoimmune processes [(95); (96)]. The 1q region is also orthologous to a region that studies suggest contains a risk factor for experimental autoimmune encephalomyelitis (EAE) in the rat [(97)].

Chromosome 5q is another region of interest from the screen that is supported by several lines of evidence. Marker D5S816 in the 5q31 region met the multipoint criterion for further interest with a Z score of 2.17. Like chromosome 1q, chromosome 5q has been investigated for a risk factor in other inflammatory and autoimmune disorders, including Crohn's disease, type 1 diabetes, celiac disease, and asthma and allergy [(98); (99); (100); (101); (102)]. Linkage to the 5q region has also been suggested in another recent multiple sclerosis screen and is orthologous to a region suggested to contain a risk factor for EAE in the rat [(44); (97)].

The initial MSGG screen on 52 families identified 19 regions that potentially harbor MS susceptibility loci [(39)]. Follow-up of these regions with an expanded data set of 96 families continued to provide the strongest support for five regions: 6p21, 6q27, 12q23-24, 16p13, and 19q13 [(48)]. Three of these five regions continue to generate HLOD scores > 1.5 in this second-generation genomic screen of 245 families (6p21, 12q24, and 16p13), providing consistent support for these regions in three of our studies to date.

Although evidence for chromosome 19q13 has been consistently seen in our families, there is decreased evidence for this region in the current study. Despite this decreased evidence, 19q13 remains interesting for several reasons. Outside of the MHC, 19q13 is the region most consistently observed for linkage and/or association with MS. At least five genomic screens, including the initial MSGG screen and the present screen (HLOD = 1.44), demonstrate at least moderate evidence of linkage to 19q13 [(39); (38); (40); (41); (45)]. Numerous allelic association studies also provide evidence of a risk locus in this

region. In addition, there is substantial evidence that the ApoE gene in this region modulates the severity and/or progression of MS [(67); (72); (68); (71); (70); (73); (74); (103)]. Further investigation will be necessary to confirm and identify a specific 19q13 disease locus.

As suggested in the literature, stratification yielded substantial increases in our linkage signals in several defined data sets [(104)]. In addition, four regions yielded HLOD scores > 2.0 only when HLA-DR2 stratification was performed: 2q was identified in the U.S. DR2- families, 6q was identified in the French DR2- families, and 13q and 16q were identified in both the U.S. DR2+ and overall DR2+ families. Results in the 2q and 6q regions suggest effects independent of HLA-DR, while results in the 13q and 16q regions suggest potential interactive effects with HLA-DR. However, formal tests of correlations between NPL scores in families with at least one affected sib-pair (n = 236) and NPL scores for regions on chromosomes 1, 3, 5, 9, and 16 failed to detect any significant correlations.

The next step of investigation is to narrow the regions of interest identified in this screen with genotyping data for a denser map of SNP markers. Several new techniques allow for rapid and accurate genotyping of large numbers of SNPs in small regions, thus promoting high levels of information extraction from a given data set. Finer mapping studies were performed using these techniques in Specific Aim 2b to further localize chromosomal regions 1q, 2q, 9q, 13q, 16q, 18p, and 19q in our MS data set. Results of these studies are presented in Chapter 5.

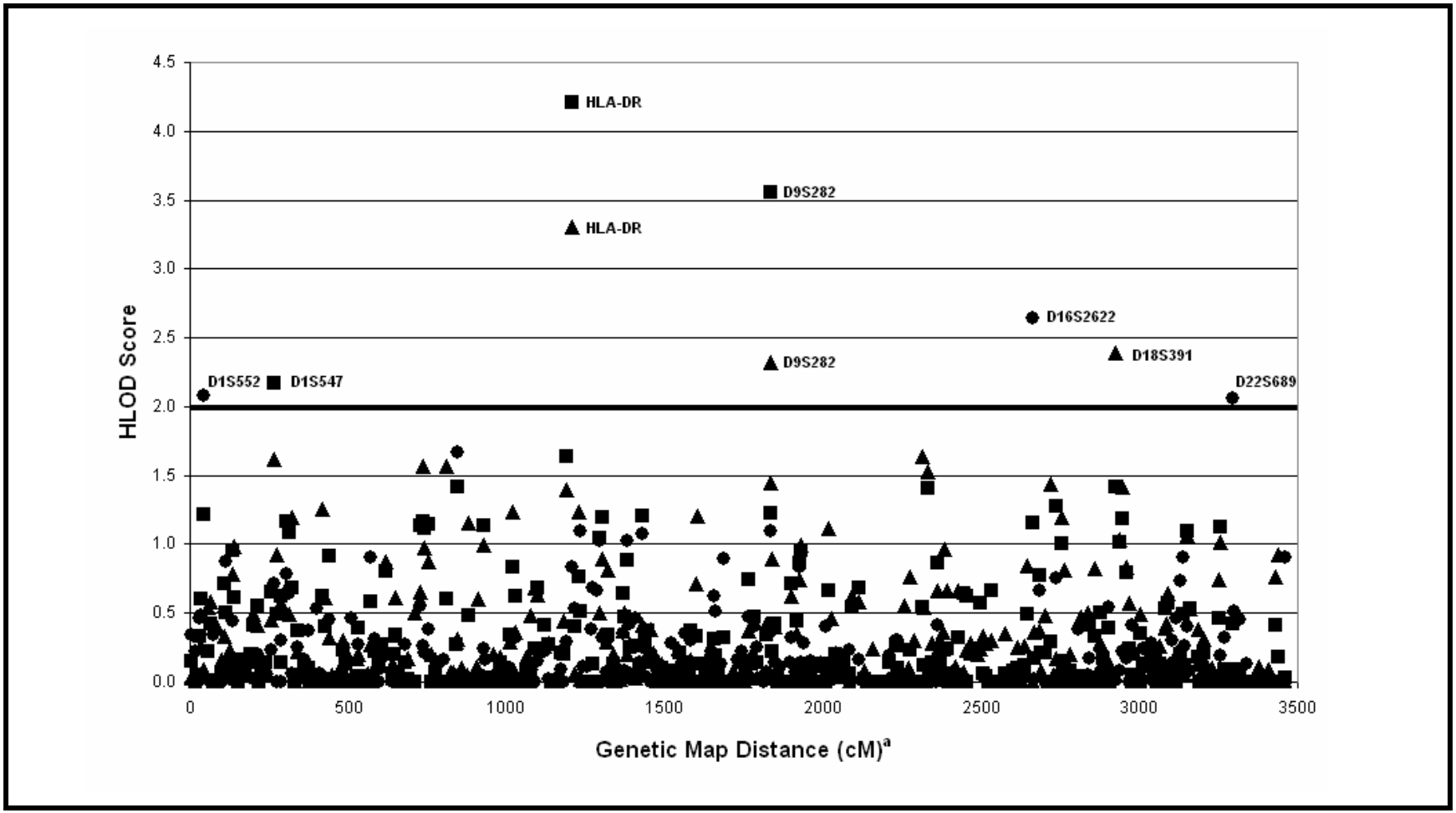


Figure 1. Two-Point HLOD Scores for the Overall Data Sets ▲ = U.S. data set (151 families); ● = French data set (94 families); ■ = combined data sets (245 families)**;** [a]Marshfield genetic map (see Marshfield website)

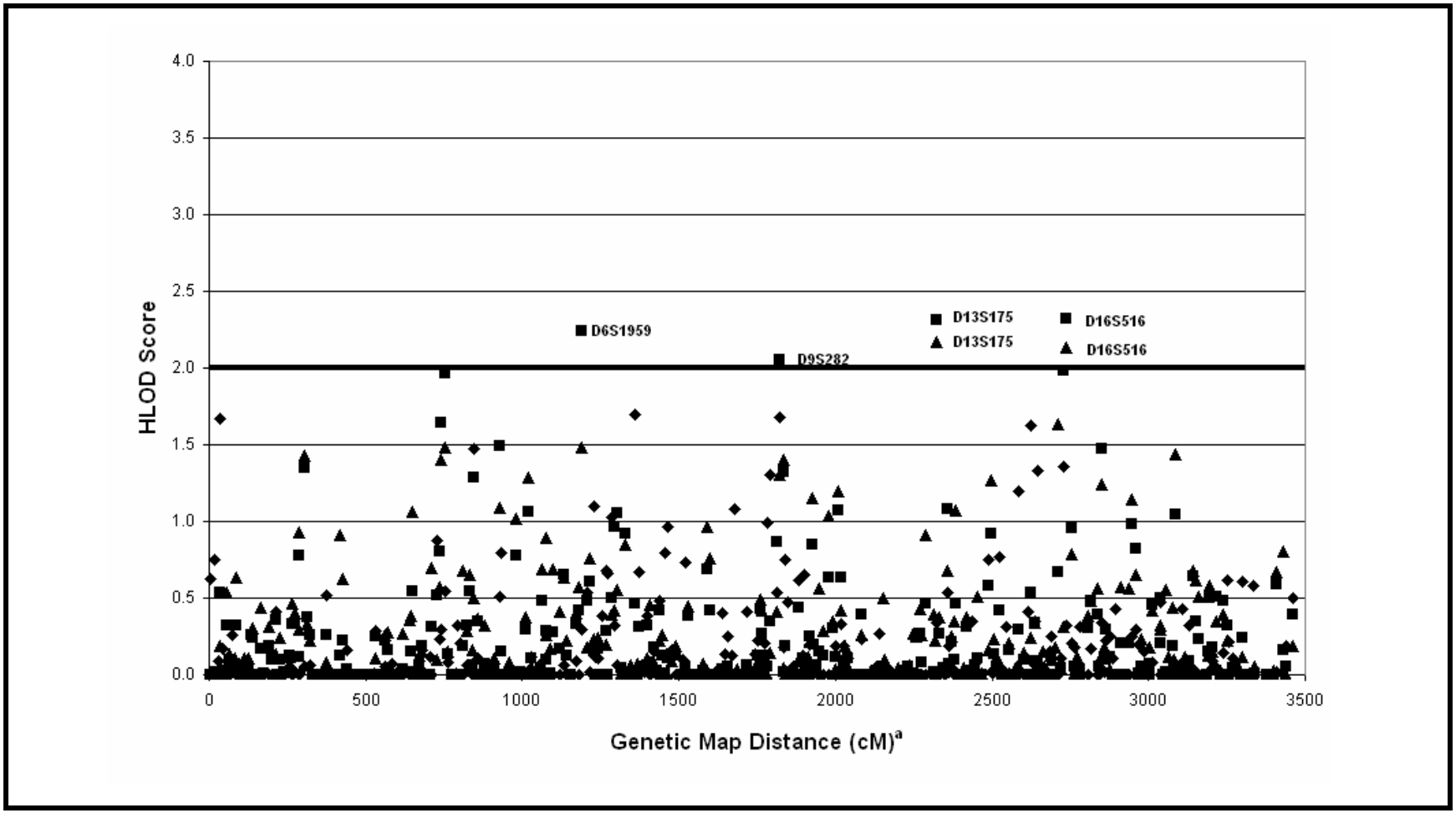


Figure 2. Two-Point HLOD Scores for the HLA-DR2+ Subsets ▲ = U.S. data set (83 families); ● = French data set (28 families); ■ = combined data sets (111 families)**;** [a]Marshfield genetic map (see Marshfield website)

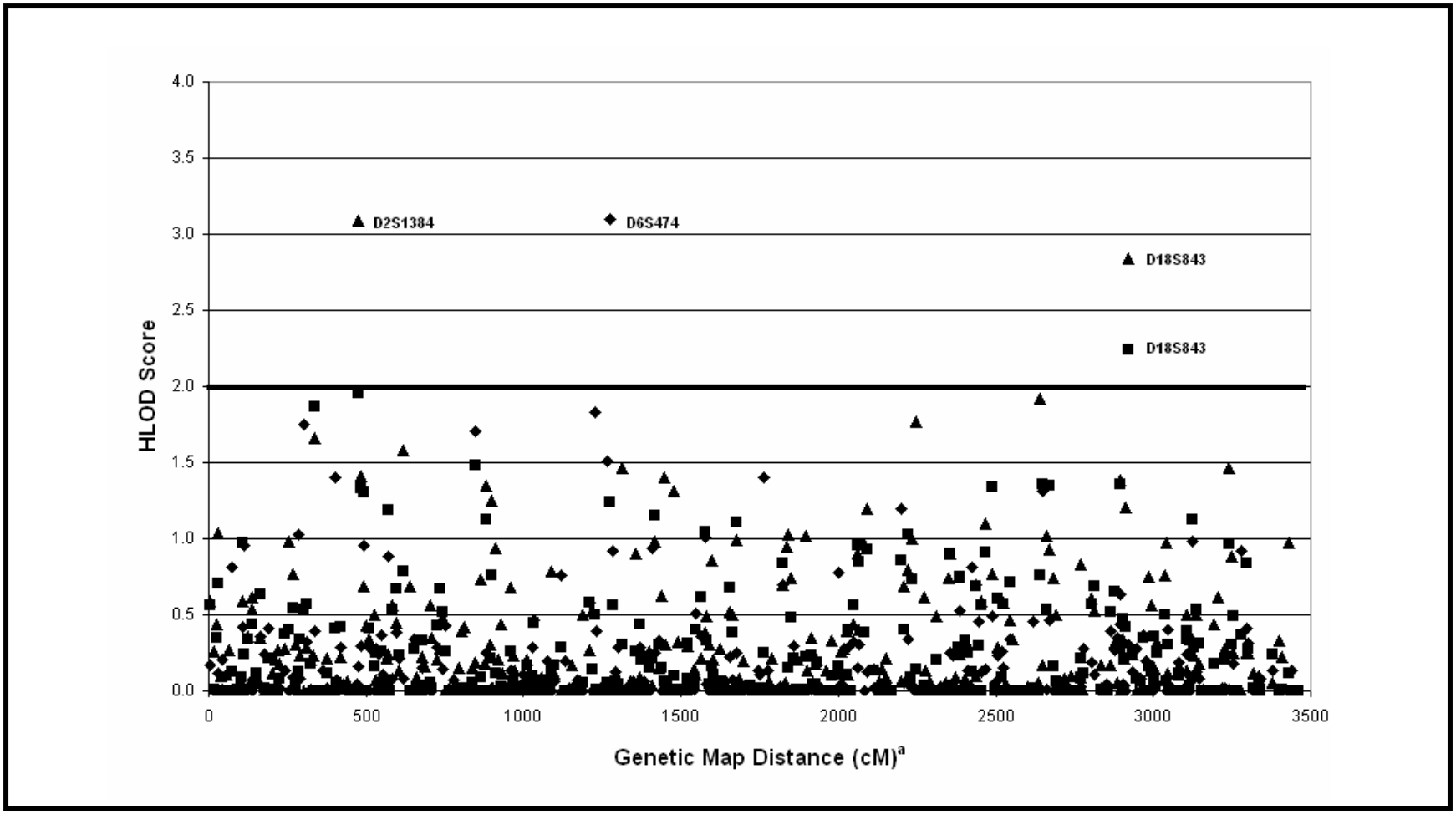


Figure 3. Two-Point HLOD Scores for the HLA-DR2- Subsets ▲ = U.S. data set (31 families); ● = French data set (35 families); ■ = combined data sets (66 families)**;** [a]Marshfield genetic map (see Marshfield website)

Table 2. Regions with Two-Point HLOD Scores > 2.0 Highest HLOD scores are indicated in **bold** text.

Chromosome	Location (cM)[a]	Marker	Data Set[b]	HLOD score
1p	45	D1S552	3	**2.08[r]**
1q	268	D1S547	1	**2.17[r]**
2q	200	D2S1384	7	**3.09[d]**
6p	34	D6S1959	4	**2.24[r]**
	44	HLA-DR	1, 2	**4.21[d]**, 3.30[r]
6q	119	D6S474	8	**3.10[r]**
9q	136	D9S282	1, 2, 4	**3.55[r]**, 2.32[r], 2.05[r]
13q	6	D13S175	4, 5	**2.32[r] ,** 2.17[r]
16p	8	D16S2622	3	**2.64[r]**
16q	100	D16S516	4, 5	**2.32[r],** 2.13[r]
18p	19	D18S391	2	**2.39[d]**
	28	D18S843	6, 7	2.25[d]**, 2.84[d]**
22q	29	D22S689	3	**2.06[r]**

[a]Marshfield genetic map (see Marshfield website)

[b]1 = overall; 2 = U.S. only; 3 = French only; 4 = all DR2+ families; 5 = U.S. DR2+ families; 6 = all DR2- families; 7 = U.S. DR2- families; 8 = French DR2-families

[d] HLOD scores calculated under a dominant model;

[r] HLOD scores calculated under a recessive model

Table 3. Regions with Multipoint Z Scores > 2.0

Chromosome	Location (cM)[a]	Marker	Z score
1q	268	D1S547	3.38
3q	216	D3S2418	2.39
5q	139	D5S816	2.17
6p	34	D6S1959	2.26
9q	136	D9S282	2.71
16p	8	D16S2622	2.05

[a]Marshfield genetic map (see Marshfield website)

CHAPTER V

EXAMINATION OF SEVEN CANDIDATE REGIONS FOR MULTIPLE SCLEROSIS: STRONG EVIDENCE OF LINKAGE TO CHROMOSOME 1Q44[§]

Abstract

Multiple sclerosis (MS) is a debilitating neuroimmunological and neurodegenerative disorder that affects ~ 2.5 million people worldwide. Genomic screens have identified numerous chromosomal regions of interest for MS loci. However, with the exception of the human leukocyte antigen (HLA) locus, studies have failed to consistently identify genes that confer disease susceptibility. An MS data set of 173 multiplex families was used to further investigate seven non-HLA regions (1q, 2q, 9q, 13q, 16q, 18p, and 19q) identified in a recent genomic screen conducted by the U.S. and French Multiple Sclerosis Genetics Groups (see Chapter 4). Single nucleotide polymorphisms (SNPs) were genotyped at ~ 1 Mb intervals extending ≥ 10 Mb to each side of peak genomic screen markers. Parametric two-point analyses identified peak HLOD scores > 2.0 for regions 1q44 (HLODs = 2.07 and 2.60) and 19q13 (HLOD = 2.01). Non-parametric multipoint analyses identified a peak LOD* score of 2.99 for the 1q44 region and substantially narrowed the linkage peak in this region to ~ 7 Mb (corresponding to a – 1.0 LOD score confidence interval). Ordered subset analyses (OSA) identified significant LOD score increases in regions 2q35 and 18p11 when ranking families by HLA-DR

[§] Chapter adapted from:

Kenealy, S.J., Herrel. L.A., Bradford, Y., Schnetz-Boutaud, N., Oksenberg, J.R., Hauser, S.L., Barcellos, L.F., Schmidt, S., Pericak-Vance. M.A., Haines, J.L. (2005) Examination of Seven Candidate Regions for Multiple Sclerosis: Strong Evidence of Linkage to Chromosome 1q44. Submitted.

status and identified a significant LOD score increase in region 2q35 when ranking families by linkage to chromosome 1q44. The increased evidence of linkage to 1q44 is particularly interesting in light of linkage evidence for this region in studies of both rheumatoid arthritis and systemic lupus erythematosus. A more detailed examination of the 1q44 region is currently underway.

Introduction

Over 70 genomic regions have been investigated in an attempt to identify MS loci. However, the genomic screens identifying these regions have lacked sufficient power to identify loci of moderate effect with statistically significant results. Although it is probable that several of the regions identified in genomic screens for MS loci actually harbor disease loci, the failure to replicate results in multiple screens or in follow-up studies is problematic.

We recently completed the largest genomic screen for MS to date. In conjunction with the French Multiple Sclerosis Genetics Group, we genotyped 361 microsatellite markers in 245 multiplex families consisting of 344 affected sib-pairs and 112 other affected relative pairs. In addition to the HLA-DR locus, the strongest signals from the screen were generated in regions 1q, 2q, 9q, 13q, 16q, and 18p (HLOD scores and/or Z scores > 2.0) [(105)] (see Chapter 4).

In the current follow-up study, we further investigated these six non-HLA regions in an expanded U.S. data set (containing 29 additional multiplex families) with increased

marker coverage. Although chromosome 19q did not meet the formal criterion for further interest in the screen, this region is frequently observed in linkage studies for MS and was also investigated. Additional marker coverage with SNPs was performed to increase information extraction and narrow the linkage signals in each follow-up region.

Material and Methods

Families

The data set used in this study consisted of U.S. families from our recently published genomic screen [(105)] and 29 newly ascertained U.S. families (containing 11 affected sib-pairs, 56 discordant sib-pairs, and 23 other affected relative pairs). The full data set contained 173 multiplex families consisting of 451 affected individuals, 251 affected sib-pairs, and 163 other affected relative pairs (Table 1).

Table 1. Description of the Data Set Families were designated HLA-DR2+ if all affected individuals carried at least one HLA-DR2 allele or designated HLA-DR2- if no affected individuals carried an HLA-DR2 allele.

# Multiplex Families	# Affecteds	# ASPs	# Other ARPs	# HLA-DR2+ Families	# HLA-DR2- Families
173	451	251	163	93	36

Families were ascertained by the University of California at San Francisco (UCSF) from across the U.S. All affected family members were examined or had their medical records

reviewed by a collaborating physician. Families were extended through all affected first-degree relatives when possible.

All protocols were approved by the appropriate Institutional Review Boards and all individuals provided informed consent before participating in the study. Positive family histories were investigated by direct contact with other family members, request for medical records, and by clinical examination, laboratory testing, or paraclincial studies (MRI scanning and evoked-response testing). Individuals were placed into one of four categories: definite MS, probable MS, possible MS, and no evidence of MS. Consistent and stringent clinical criteria were applied as described elsewhere [(5); (39)] and all clinically definite MS cases met the Poser criteria [(106)]. Only definite MS cases were classified as affected individuals in the analyses.

Molecular Analysis

After obtaining informed consent, blood samples were collected from each study participant. Genomic DNA was extracted from blood samples using standard procedures as described elsewhere [(89)]. All DNA samples were coded and stored at $4^{o}C$ prior to use.

The Celera and Applied Biosystems databases were mined to select SNPs according to location relative to other selected SNPs, high minor allele frequency ($\geq$ 0.40), and availability of quality assays (see Celera website; ABI website). SNPs were genotyped at ~ 1 Mb intervals extending $\geq$10 Mb to each side of peak markers from the genomic

screen. Additional SNPs were genotyped if linkage curves were unresolved at the edge of any follow-up region (for example, 15 additional SNPs were genotyped to resolve OSA peaks generated for the HLA-DR LOD score Low to High and chromosome 1 LOD score High to Low ranking schemes on chromosome 2q35). SNP genotyping was performed using Assays-On-Demand^{SM} or Assays-by-Design^{SM} with the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). Amplification was performed in a 384-well GeneAmp® PCR System 9700 (Applied Biosystems, Foster City, CA) or a 384-well DNA Engine Tetrad® 2 Peltier Thermal Cycler (MJ Research, Watertown, MA) with the following conditions: 94°C - 10 min; 92°C - 15 sec, 60°C - 1 min (50 cycles); 4°C - hold. Genotypes for HLA-DR were determined using non-radioactive PCR-SSOP (Dynal Biotech, Brown Deer, WI).

Marker order and intermarker distance were obtained from Celera reference maps and verified in dbSNP build 124 (see Celera website; NCBI dbSNP website). Laboratory personnel were blinded to pedigree structure, affected status, and location of quality control samples (3 unblinded CEPH individuals and 4 blinded controls). Duplicate quality control samples were placed both within and across plates and equivalent genotypes were required to ensure accurate genotyping.

Both affected and unaffected individuals were genotyped to maximize marker information content. Information content was determined using Allegro [(93)]. Allele frequencies were calculated from the genotyped founders in each family. Hardy-Weinberg calculations were performed for each marker and Mendelian inconsistencies

were identified using PedCheck [(90)]. Improbable double recombination events were identified using SIMWALK version 2.9 [(88)]. Suspect genotypes were re-read and/or re-run. All SNPs were required to have > 90% of possible genotypes. Verification of relationships between pairs of samples within families was performed using RELPAIR [(91)]. Markers and samples failing to pass quality control measures were dropped from the analyses.

Statistical Analysis

Both model-based and model-free statistical analyses were performed. Model-based analyses were conducted using autosomal dominant and autosomal recessive models with disease allele frequencies of 0.01 and 0.20 (respectively) to model a common susceptibility allele. A penetrance value of 0.95 was used for both dominant and recessive models and individuals with no evidence of MS were coded as normal individuals for these analyses. Two-point LOD scores were calculated in FASTLINK [(84); (85)] and two-point heterogeneity LOD (HLOD) scores were calculated in HOMOG [(86); (87)].

Multipoint LOD scores for the overall data set, HLA subsets, and conditional analyses were calculated in Allegro and are reported as LOD* scores [(93)]. Optimal OSA subsets were identified using OSA software [(107)]. Corresponding LOD scores for OSA subsets were calculated in GENEHUNTER-PLUS and are reported as maxLOD scores [(108); (92)]. Statistically significant increases in LOD scores for OSA subsets were

identified by p values < 0.05. A complete list of SNP markers and individual LOD scores is available online (see CHGR Supplemental Data website).

Covariate Analysis

To account for potential locus heterogeneity, analyses were performed incorporating HLA-DR status. In subset analyses, families were selected and designated HLA-DR2+ if all affected individuals carried at least one HLA-DR2 allele (93 families) or HLA-DR2- if no affected individuals carried an HLA-DR2 allele (36 families) (Table 1).

In conditional analyses, two HLA-DR weighting schemes were used to incorporate information arising from linkage and association**.** In the first scheme, positive HLA-DR LOD scores were used to weight families, while families generating negative HLA-DR LOD scores were given a weight of zero (HLA LOD Weighted). Conversely, |negative| HLA-DR LOD scores were used to weight families, while families generating positive HLA-DR LOD scores were given a weight of zero (HLA LOD Inverse Weighted). In the second scheme, each family was given a nominal weight (between 0.0 and 1.0) to account for the number of HLA-DR2 alleles in affected individuals (HLA-DR2 Allele Weighted) or an inverse nominal weight (1 - [HLA-DR2 weight]) to account for the absence of HLA-DR2 alleles in affected individuals (HLA-DR2 Allele Inverse Weighted).

Ordered subset analysis (OSA) was performed to identify homogenous subsets of families contributing maximally to linkage in each chromosomal region [(107)]. The

OSA method was developed to reduce sample heterogeneity (for increased power to detect linkage), generate more distinctive LOD score peaks, and define maximally informative data sets for regions of interest. OSA rank orders families by a trait-related covariate (e.g. family-specific LOD scores at a second locus) to identify a subset generating maximally increased evidence of linkage relative to the overall sample. Statistical significance for this increased evidence of linkage is assessed using a random permutation procedure to estimate empirical p values. Families were ranked by family-specific LOD scores for HLA-DR (Low to High or High to Low) and family-specific HLA-DR2 allele weights (Low to High or High to Low). Because chromosome 1 demonstrated compelling evidence of linkage after follow-up genotyping and analysis, OSA was also performed for chromosomes 2q, 9q, 13q, 16q, 18p, and 19q with families ranked by chromosome 1 LOD* scores (Low to High or High to Low). OSA analysis was also performed for chromosome 1q with families ranked by chromosome 2 LOD* scores at the peak location for the chromosome 1 LOD* ranked subset (Low to High or High to Low).

Results

A list of peak two-point, multipoint, and HLA conditional LOD scores for each follow-up region is provided in Table 2. A complete report of linkage analyses for all follow-up regions is provided in Appendix C.

Chromosome 1q44 demonstrated the strongest evidence of linkage after follow-up, with SNPs replicating linkage peaks < 1 Mb from the peak genomic screen marker for the

region (D1S547). Two-point analysis generated a peak HLOD score of 2.60 in the overall data set—greater than the peak HLOD score observed in the genomic screen. Multipoint analysis generated a peak multipoint LOD* score of 2.99 in the overall data set and substantially narrowed the interval of interest in this region (Figure 1). Analyses conditioning on HLA-DR2 allele weights and HLA-DR LOD scores continued to support linkage to this region, but failed to identify a specific HLA-DR subset demonstrating evidence of interaction with the 1q44 region. Analysis of OSA subsets generated peak LOD scores greater than multipoint scores for the overall data set and for the HLA-DR covariate approaches, with OSA maxLOD scores ranging from 3.47 to 4.54. However, OSA analysis also failed to differentiate specific subsets accounting for the genetic effect in this region, as no LOD score increases were statistically significant.

Two-point analysis identified five markers in the 2q35 region with HLOD scores > 1.0, all of which were identified in the HLA-DR2- subset. While chromosome 2q35 did not demonstrate substantial evidence of linkage in multipoint calculations for the overall data set or HLA conditional analyses, OSA analysis generated a significantly increased maxLOD score of 1.86 at ~ 194 Mb in the HLA-DR LOD score Low to High ranking (15 families; $p = 0.03$). The two-point and OSA analysis subsets are similar to the U.S. HLA DR2- subset that identified linkage to 2q35 in the genomic screen. OSA analysis also generated a significantly increased maxLOD score of 2.20 at ~ 191 Mb in the chromosome 1 LOD* score High to Low ranking (22 families; $p = 0.01$) (Figure 2).

Chromosome 9q34 demonstrated modest evidence of linkage in two-point, multipoint, and HLA conditional analyses. OSA analysis increased the peak maxLOD score for this region in the HLA-DR LOD Low to High ranking, but with only a trend toward significance (maxLOD = 1.81; 91 families; p = 0.09).

Chromosome 13q11 failed to demonstrate any evidence of linkage in the multipoint or HLA conditional analyses. However, the highest two-point (HLOD = 1.36) and OSA (maxLOD = 1.08 in the chromosome 1 LOD* Low to High subset; p = 0.09) LOD scores for this region were generated adjacent to the peak microsatellite marker for 13q11 in the genomic screen.

Chromosome 16q23 generated a peak multipoint LOD* score of 0.51 in the overall data set, with a slight increase in the HLA-DR2+ subset (LOD* = 1.00; 93 families) and decrease in the HLA-DR2- subset (LOD* = 0.35; 36 families). However, conditional analysis generated the highest LOD* scores in the HLA-LOD Negative (LOD* = 1.00) and HLA-DR2 Inverse Weight (LOD* = 0.61) subsets—raising questions about the origin of the signal in this region. OSA analysis generated a peak maxLOD score of 1.83 in the chromosome 1 LOD* High to Low ranking (34 families; p = 0.14), but did not significantly increase LOD scores or differentiate linkage contribution from any of the optimal HLA-DR subsets.

Two-point analysis identified 2 markers in the 18p11 region with HLOD scores > 1.0, both of which were identified in the HLA-DR2- subset. Multipoint analysis generated a

peak LOD* score of 0.75 in the overall data set, with the HLA DR2- subset generating a slightly higher peak (LOD* = 1.01). Conditional analyses demonstrated the highest linkage scores in the HLA-DR2 Inverse Weight subset (LOD* = 1.21), while the OSA analysis demonstrated a statistically significant increase in evidence of linkage with a maxLOD score of 1.91 in the HLA-DR2- Weight Low to High ranked families (165 families; p = 0.02) (Figure 3). These two-point, multipoint, and conditional analyses provide consistent evidence of linkage when accounting for the absence of HLA-DR effects, and are also consistent with the U.S. HLA-DR2- subset that identified linkage to 18p11 in the genomic screen. However, despite a general trend of linkage in families without evidence for HLA-DR contribution, no consistent peak location was identified for this region. OSA also generated an increased LOD score with a trend toward significance in the chromosome 1 LOD* Low to High subset (maxLOD = 2.14; 126 families; p = 0.09) (Figure 3).

Chromosome 19q13 generated peak multipoint scores at the same Mb location for the overall and HLA-DR2+ subsets (LOD* = 0.86 and LOD* = 0.85, respectively). HLA conditional analyses generated a similar peak score in the HLA-DR Inverse Weight subset (LOD* = 0.85). However, OSA analyses failed to generate even moderate p values for any of the ranking strategies.

Discussion

Traditional follow-up studies of genomic screen results entail genotyping additional microsatellite markers located at smaller genetic intervals than the original screen

markers. While this approach can be effective, we chose to follow up our genomic screen results with a denser map of SNPs. With higher throughput capabilities and more accurate genotypes, SNPs allowed for expedited genotyping and improved quality control of our data. As suggested by a recent publication, increased information content from these markers also provided more thorough coverage of each region and substantially increased the effective size of our data set by 35% (Figure 4) [(109)]. Follow-up with SNPs located at ~ 1 Mb intervals resulted in higher quality data, greater information extraction, and substantially narrowed linkage peaks in several of our regions of interest.

Although the linkage peaks in several of our follow-up regions were generated near linkage peaks from the genomic screen, it is important to consider the potential effects of using approximate marker locations in this study. Given the inaccuracy of interpolating genetic distances between SNPs using currently available linkage maps, our study design used physical distances as an approximation of genetic distances in the multipoint calculations. The marker order for each region was verified in multiple databases and the reported recombination rates for these regions did not substantially deviate from the 1cM/1Mb paradigm (see deCODE website). While we therefore determined that using physical distances was most appropriate for the current study design, it is important to recognize this underlying assumption concerning genetic distances when assessing the multipoint linkage results.

The strongest linkage result from this study arises on chromosome 1q44, where the multipoint LOD score increased relative to the initial genomic screen. In addition to the

present study, several lines of evidence support chromosome 1q. Marker D1S547 met the criterion for further interest in both two-point and multipoint calculations in our recent second-generation genomic screen (HLOD = 2.17; Z = 3.38) [(105)] (see Chapter 4). Nearby markers have also demonstrated suggestive linkage and/or association in several other MS screens that have been conducted in a variety of study populations [(44); (47); (52); (53); (94)]. Another compelling piece of evidence for 1q is linkage to this region in studies for the autoimmune diseases rheumatoid arthritis [(95); (110); (111)] and systemic lupus erythematosus [(112); (113); (96); (114)], suggesting the presence of a gene for general autoimmune processes in this region. Interestingly, all of these markers demonstrating linkage in the MS, rheumatoid arthritis, and systemic lupus erythematosus screens are within 6 Mb of the peak marker from our genomic screen and follow-up study.

Our follow-up also provides evidence that covariate analysis can be a useful approach for identifying subsets, increasing evidence of linkage, and narrowing confidence intervals in regions of interest. Occasional differences in subsets and peak locations identified by the conditional analyses in this study indicate that these schemes may vary in their power to capture moderate genetic effects or are measuring slightly different biological phenomena. However, the overall consistency suggests that additional information gained from conditional analyses may be valuable in future study designs.

The OSA method proved to be especially useful in regions failing to demonstrate evidence of linkage in the overall data set. Although OSA p value thresholds were not

corrected for multiple testing, this method provided additional evidence for linkage and aided in the identification and interpretation of multilocus effects. The potential for OSA was initially illustrated with breast cancer data, where linkage to the chromosome 17q region containing BRCA1 was identified when ranking families by mean age of onset [(107)]. The OSA method has subsequently been applied to data sets for Alzheimer disease, autism, age-related macular degeneration, type 2 diabetes, and prostate cancer [(115); (116); (117); (118); (119); (120); (121)]. Modification of OSA and other covariate methods to account for multiple loci or other relevant factors would further strengthen these approaches. Further elucidation of the genetic effects in known regions of linkage such as HLA-DR and chromosome 1 could also serve as crucial tools in unraveling the genetic heterogeneity of MS.

This study confirmed several chromosomal regions warranting further investigation in the search for genes conferring susceptibility to MS. Due to the continued evidence of linkage and the consistency of data sets from which these linkage signals arise, we are particularly interested in pursuing chromosome regions 1q44, 2q35, 9q34, and 18p11. A more detailed investigation of chromosome 9q34 is presented in Chapter 6. A more detailed investigation of the ~ 7 Mb interval on chromosome 1q44 is presented in Chapter 7.

Table 2. Peak LOD Scores

Chromosome Region	Peak Screen HLOD Score (Kenealy et al. 2004)	Peak Follow-Up HLOD Score	Peak Follow-Up Multipoint LOD* Score	Peak Follow-Up HLA Conditional LOD Score
1q44	2.17	2.60	2.99	3.01[b]
2q35	3.09	0.82	0.00	0.59[b]
9q34	3.55	1.89	0.63	0.93[d]
13q11	2.32	0.92	0.01	0.14[e]
16q23	2.32	1.25	0.51	1.00[f]
18p11	2.84	0.93	0.75	1.21[d]
19q13	1.44	2.01	0.85	0.86[a]

Data Sets: a = HLA-DR2+ (93 Families); b = HLA-DR2- (36 Families); c = HLA-DR2 Allele Weighted (173 Families); d = HLA-DR2 Allele Inverse Weighted (173 Families); e = HLA LOD Weighted (173 Families); f = HLA LOD Inverse Weighted (173 Families)

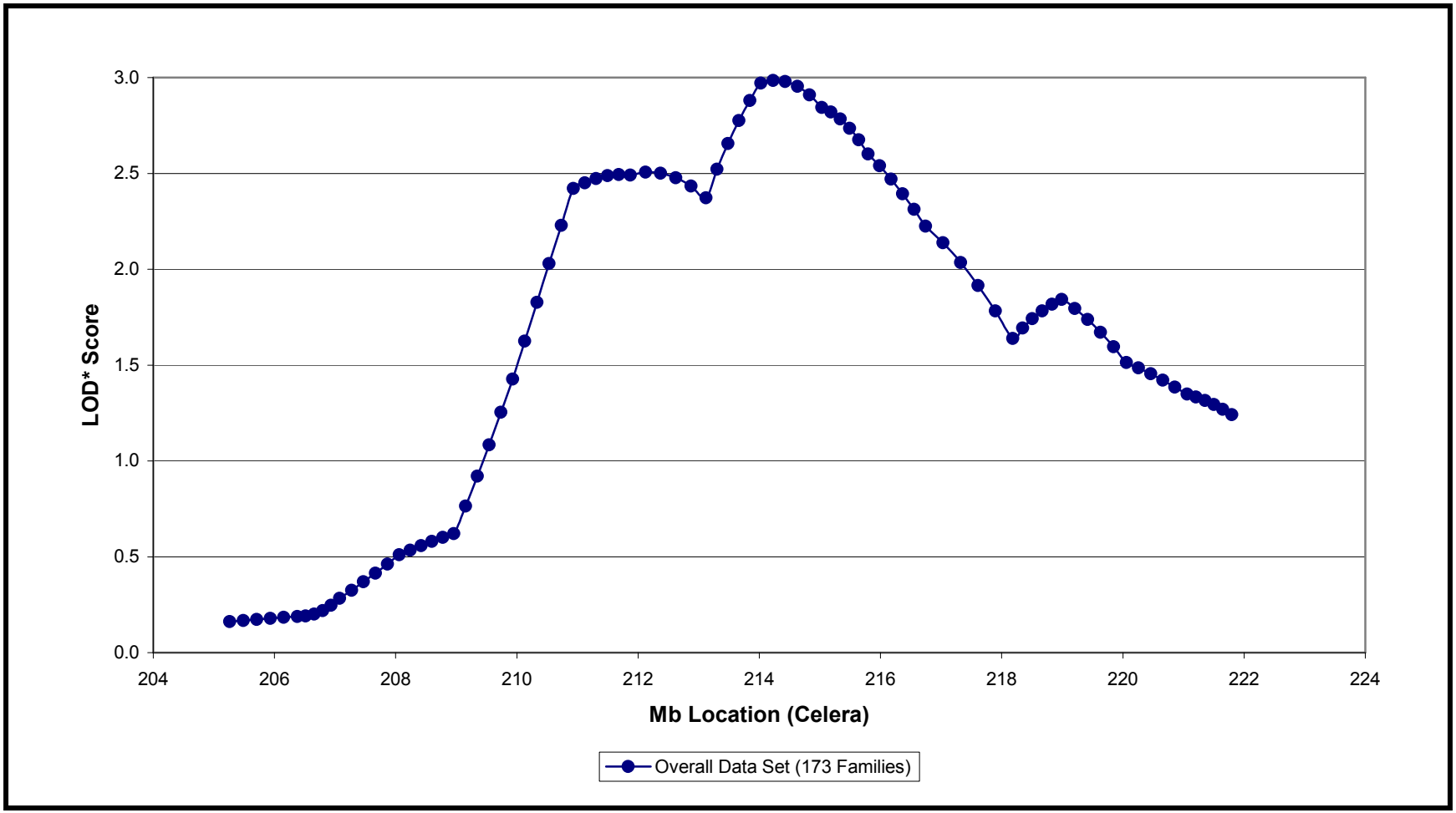


Figure 1. Chromosome 1q Multipoint Analysis

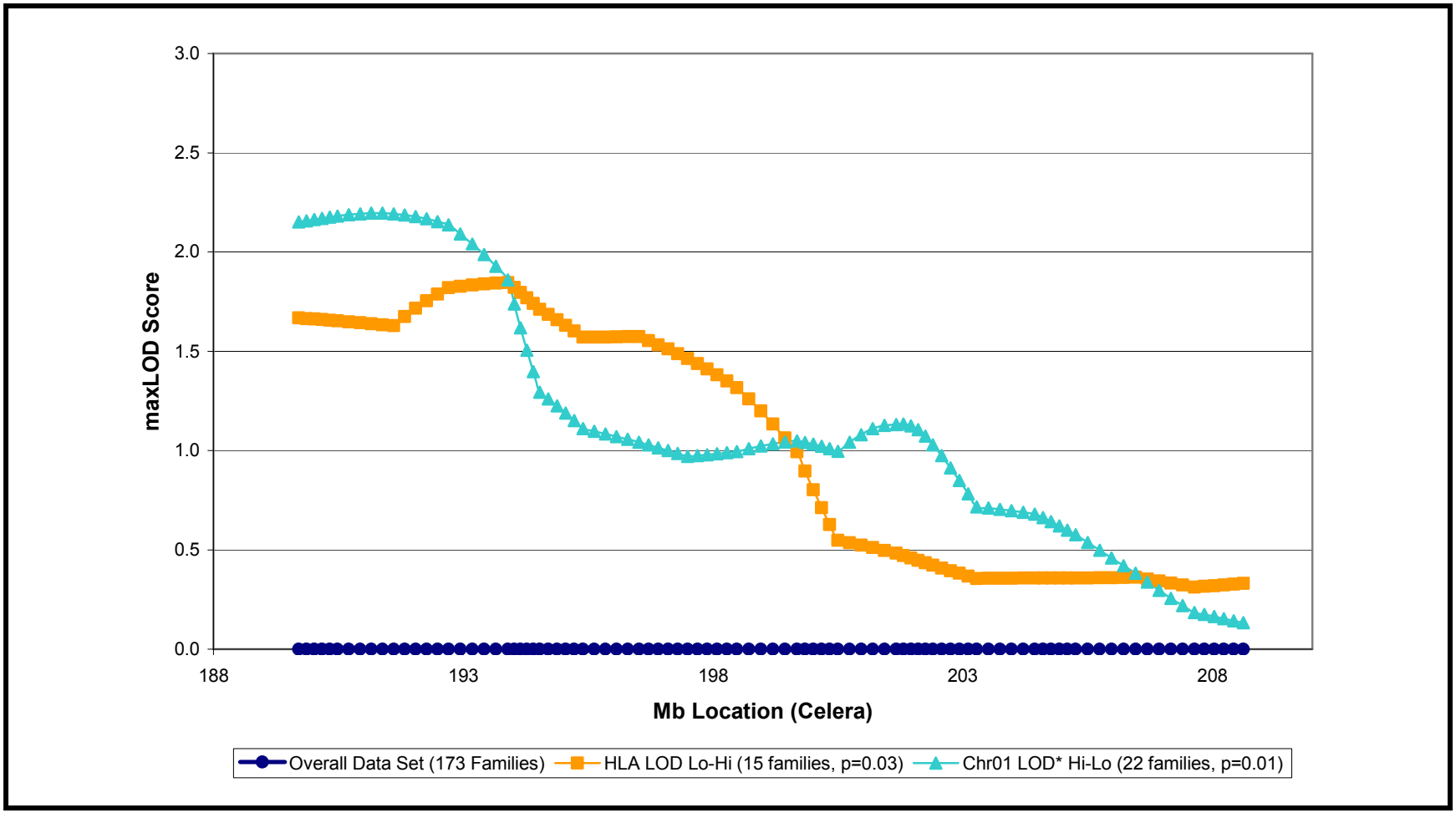


Figure 2. Chromosome 2q OSA Analysis

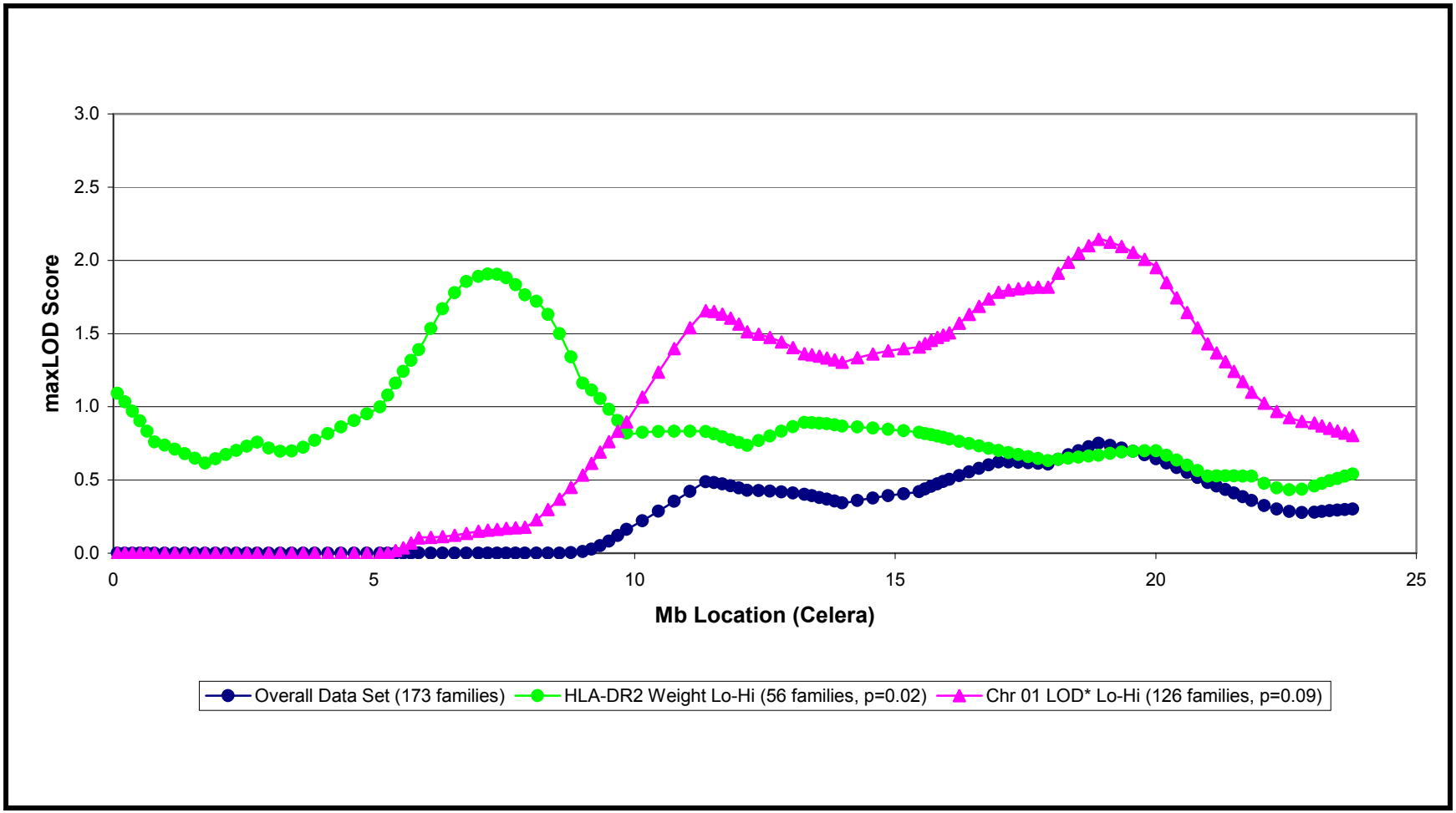


Figure 3. Chromosome 18p OSA Analysis

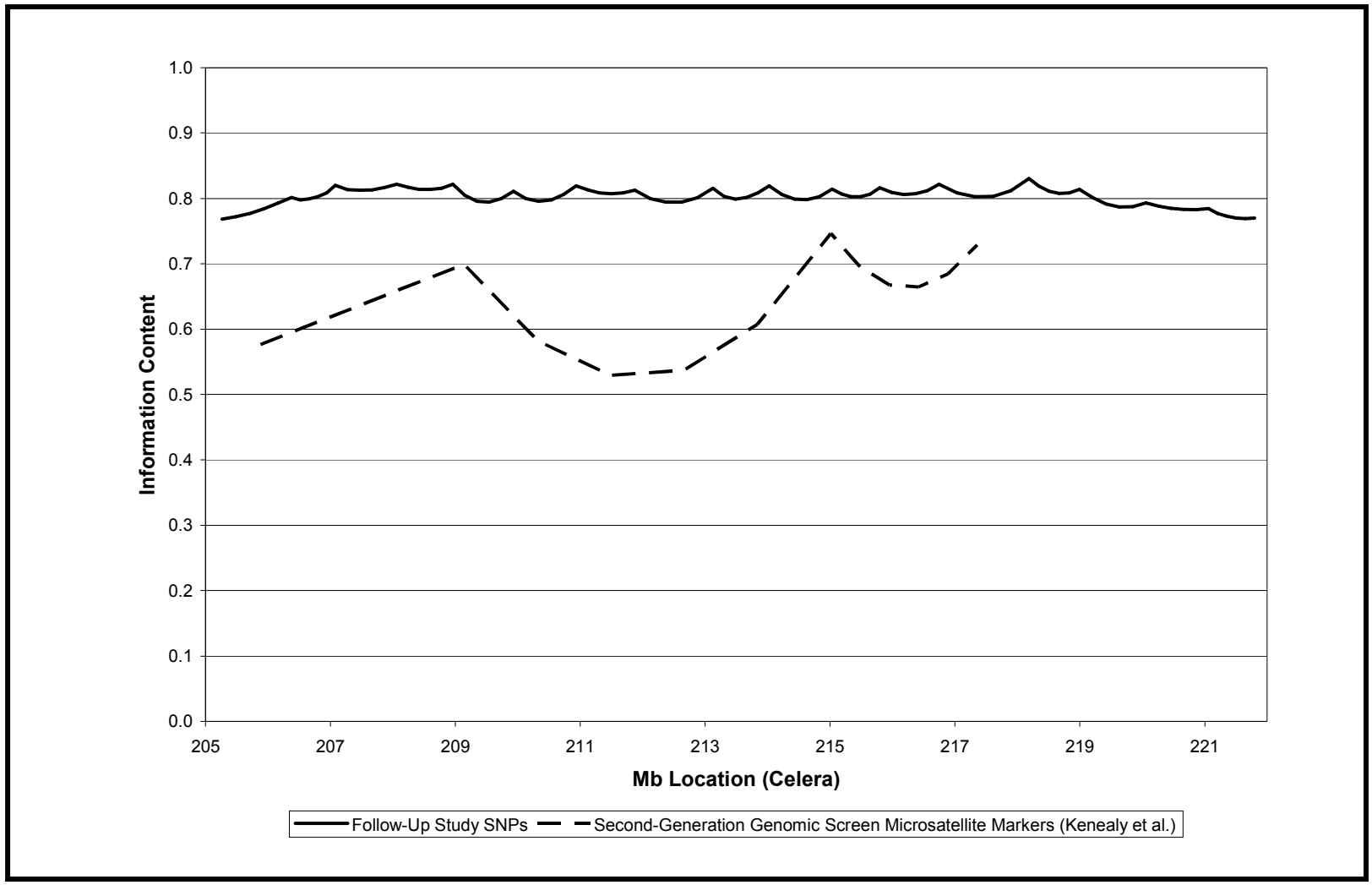


Figure 4. Information Content for Markers Genotyped on Chromosome 1q

CHAPTER VI

A DIRECTED INVESTIGATION OF CHROMOSOME 9Q34

Introduction

Several of our studies suggest the presence of a gene for MS susceptibility in the chromosome 9q34 region. The initial Multiple Sclerosis Genetics Group (MSGG) genomic screen and follow-up study detected moderate evidence of linkage to 9q34 in two-point LOD score analyses (HLOD = 1.13; HLOD = 1.40) [(39); (48)]. In the second-generation MSGG genomic screen, 9q34 demonstrated the strongest evidence of linkage to a non-MHC region in the overall data set (HLOD = 3.55; Z = 2.71) (see Chapter 4). The follow-up study to the second-generation genomic screen continued to provide evidence of linkage to 9q34, with moderate evidence of linkage in two-point, multipoint, and HLA conditional analyses. OSA increased evidence of linkage to this region in the HLA-DR LOD Low to High ranking with a trend toward significance (maxLOD = 1.81; 91 families; p = 0.09) (see Chapter 5). Recent genomic screens conducted by other research groups in Nordic sib-pairs and Turkish families also demonstrate moderate support for 9q34, highlighting the need for further investigation of this region [(46); (49)].

In addition to generating consistent evidence of linkage, the follow-up study to the second-generation genomic screen also revealed consistent location of linkage peaks at ~

106.5 Mb. Because the 9q34 linkage peak was localized and narrow, we further investigated the region located under the peak for linkage to and allelic association with additional SNP marker coverage.

Material and Methods

Families

The data set used in this detailed investigation of chromosome 9q34 included 173 multiplex families consisting of 451 affected individuals, 251 affected sib-pairs, and 163 other affected relative pairs (see Chapter 5).

Molecular Analysis

The Celera and Applied Biosystems databases were mined to select SNPs according to location relative to other selected SNPs, high minor allele frequency ($\geq$ 0.40), and availability of quality assays (see Celera website; ABI website). An additional 35 SNPs were genotyped under the linkage peak to attain coverage of ~ 1 SNP/30 Kb (Table 1). SNP genotyping was performed using Assays-On-Demand^{SM} or Assays-by-Design^{SM} with the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). Amplification was performed in a 384-well GeneAmp® PCR System 9700 (Applied Biosystems, Foster City, CA) or a 384-well DNA Engine Tetrad® 2 Peltier Thermal Cycler (MJ Research, Watertown, MA) with the following conditions: 94°C - 10 min; 92°C - 15 sec, 60°C - 1 min (50 cycles); 4°C - hold.

Marker order and intermarker distance were obtained from Celera reference maps and verified in dbSNP build 124 (see Celera website; NCBI dbSNP website). Laboratory personnel were blinded to pedigree structure, affected status, and location of quality control samples (3 unblinded CEPH individuals and 4 blinded controls). Duplicate quality control samples were placed both within and across plates and equivalent genotypes were required to ensure accurate genotyping.

Both affected and unaffected individuals were genotyped to maximize marker information content. Information content was determined using Allegro [(93)]. Allele frequencies were calculated from the genotyped founders in each family. Hardy-Weinberg calculations were performed for each marker and Mendelian inconsistencies were identified using PedCheck [(90)]. Suspect genotypes were re-read and/or re-run. All SNPs were required to have > 90% of possible genotypes. Markers and samples failing to pass quality control measures were dropped from the analyses.

Statistical Analysis

Both model-based and model-free analyses were performed. Model-based analyses were conducted using autosomal dominant and autosomal recessive models with disease allele frequencies of 0.01 and 0.20 (respectively) to model a common susceptibility allele. A penetrance value of 0.95 was used for both dominant and recessive models and individuals with no evidence of MS were coded as normal individuals for these analyses. Two-point LOD scores were calculated in FASTLINK [(84); (85)] and two-point heterogeneity LOD (HLOD) scores were calculated in HOMOG [(86); (87)].

Linkage disequilibrium (LD) patterns were constructed using Haploview software (see Haploview website). Computation of LD statistics were performed for all marker pairs and haplotype blocks were identified through pairwise r^2 values (Figure 1; Figure 2). SNPs identified as haplotype tagging for LD blocks (htSNPs) were selected for multipoint calculations. In regions where no SNP (or more than one SNP) was identified as an htSNP, SNPs with the highest observed heterozygosity and/or minor allele frequency were selected. Nineteen SNPs were removed from the data set to eliminate pairwise r^2 values > 0.10 prior to conducting multipoint linkage and association analyses (Figure 3).

Multipoint LOD scores for the overall data set, HLA subsets, and conditional analyses were calculated in Allegro and are reported as LOD* scores [(93)]. Optimal OSA subsets were identified using OSA software [(107)]. Corresponding LOD scores for OSA subsets were calculated in GENEHUNTER-PLUS and are reported as maxLOD scores [(108); (92)]. Statistically significant increases in LOD scores for OSA subsets were identified by p values < 0.05. Association analyses were performed with PDT and Haploview [(81); see Haploview website]. A complete list of SNP markers and individual LOD scores is available online (see CHGR Supplemental Data website).

Covariate Analysis

In subset analyses, families were selected and designated HLA-DR2+ if all affected individuals carried at least one HLA-DR2 allele (93 families) or HLA-DR2- if no affected individuals carried an HLA-DR2 allele (36 families) (Table 1).

In conditional analyses, two HLA-DR weighting schemes were used to incorporate information arising from linkage and association. In the first scheme, positive HLA-DR LOD scores were used to weight families, while families generating negative HLA-DR LOD scores were given a weight of zero (HLA LOD Weighted). Conversely, |negative| HLA-DR LOD scores were used to weight families, while families generating positive HLA-DR LOD scores were given a weight of zero (HLA LOD Inverse Weighted). In the second scheme, each family was given a nominal weight (between 0.0 and 1.0) to account for the number of HLA-DR2 alleles in affected individuals (HLA-DR2 Allele Weighted) or an inverse nominal weight (1 - [HLA-DR2 weight]) to account for the absence of HLA-DR2 alleles in affected individuals (HLA-DR2 Allele Inverse Weighted).

Ordered subset analysis (OSA) was performed to identify homogenous subsets of families contributing maximally to linkage in each chromosomal region [(107)]. Families were ranked by family-specific LOD scores for HLA-DR (Low to High or High to Low) and family-specific HLA-DR2 allele weights (Low to High or High to Low). Because chromosome 1 demonstrated compelling evidence of linkage after follow-up genotyping and analysis, OSA was also performed with families ranked by chromosome 1 LOD* scores (Low to High or High to Low).

Results

Two-point linkage and PDT association analysis results for all SNPs are provided in Table 1. Of the 35 new SNPs, 13 SNPs generated HLOD scores > 1.0 in the overall data

set. With the exception of microsatellite D9S2157, the highest two-point LOD score for the region was generated by a SNP located at ~ 107 Mb (HLOD = 2.73). Seven of the new SNPs generated HLOD scores > 1.0 in the HLA-DR2+ families (six of which also generated HLOD scores > 1.0 in the overall data set), while only one new SNP generated an HLOD score > 1.0 in the HLA-DR2- families. Four new SNPs generated PDT p values < 0.05, with the smallest p value being generated by a SNP also located at ~ 107 Mb (p value = 0.006).

Because Haploview revealed that several SNPs were in LD and therefore threatened to artificially inflate multipoint linkage and association results, 19 SNPs were eliminated from the data set prior to conducting multipoint analyses (Figure 1; Figure 3). These 19 SNPs represented the minimum possible number of markers that could be removed to eliminate all pairwise r^2 values > 0.10. When possible, these eliminated SNPs also represented the least informative marker in each pairwise combination.

Multipoint analysis results are provided in Figures 4-6. The addition of 16 SNPs increased the maximum LOD* score in the overall data set from 0.63 to 2.16. The maximum LOD* scores in the HLA-DR2+ was also increased, from 0.16 to 1.41. In the conditional analysis, the HLA-DR2 Allele Inverse Weighted scheme increased the maximum LOD* score from 0.93 to 2.89. And while OSA increased maxLOD scores for all ranking schemes, only ranking of families by HLA-DR2 Weight from Low to High generated a p value < 0.05 (maxLOD = 2.48; p = 0.04).

Discussion

Additional marker coverage under the 9q34 linkage peak increased evidence of a MS risk factor in both two-point and multipoint analyses. Although covariate analyses continued to yield results in data sets reflecting both the presence and absence of HLA effects, the consistent increase in evidence of linkage is encouraging. The ~ 1 Mb region investigated with increased marker coverage contains several interesting genes, including tuberous sclerosis 1 (TSC1), ABO blood group (ABO), and dopamine beta-hydroxylase (DBH) (see Celera website). Further investigation of this narrow region of interest is currently underway.

Table 1. Chromosome 9q Analysis Results Microsatellites are indicated in yellow. New SNPs are indicated in grey. Markers generating LOD scores > 1.0 are indicated in bold text.

SNP ID (Celera)	SNP ID (dbSNP)	Mb (Celera)	Mb (dbSNP)	Gene	MAF	All	HLA All	HLA None	PDTSum	PDTAve	PDTGeno
hCV1899341	rs3810928	87.463072	113.895331	AMBP	0.48	0.12	0.00	0.00	0.294	0.336	0.490
hCV3022586	rs1061494	88.494049	114.926134	TNC	0.42	0.00	0.02	0.00	0.390	0.703	0.152
hCV7593836	rs971037	89.314153	115.746096	EST-YD1	0.47	0.00	0.00	0.00	0.807	0.152	0.786
hCV1979634	rs10733620	90.172104	116.604046	ASTN2	0.44	0.00	0.05	0.00	0.615	0.353	0.625
hCV11722141	rs1927911	91.117821	117.549608	TLR4	0.24	0.72	1.00	0.09	0.223	0.292	0.431
D9S934		91.743400			9 alleles	0.21	0.15	0.07	N/A	N/A	N/A
hCV1920588	rs1324623	92.767997	119.199751	DBCCR1	0.31	0.14	0.03	0.79	0.541	0.974	0.728
hCV1219009	rs3747850	94.713428	121.144482	GSN	0.49	0.01	0.00	0.02	0.115	0.134	0.208
hCV11884087	rs4679	95.562902	121.994167	NDUFA8	0.43	0.25	0.14	0.21	0.184	0.889	0.422
hCV58657	rs12686320	96.805159	123.236485	KIAA1608	0.37	0.13	0.10	0.20	0.910	0.423	0.520
hCV8780788	rs1042486	97.430687	123.862987	LHX2	0.44	0.00	0.00	0.02	0.430	0.428	0.643
D9S282		97.456102			9 alleles	0.39	0.05	0.36	N/A	N/A	N/A
hCV302240	rs501963	98.610850	125.042470	RAB9P40	0.43	**1.94**	0.31	**1.47**	0.066	0.736	0.109
hCV8782473	N/A	99.379990	125.805096	PBX3	0.30	0.00	0.11	0.01	0.268	0.249	0.453
hCV2700890	rs874799	100.376610	126.801559	RALGPS1A	0.46	0.42	0.10	0.00	1.000	0.937	0.151
hCV580692	rs514024	101.154940	127.583624	SH2D3C	0.40	0.00	0.00	0.00	0.948	0.831	0.193
hCV16180096	rs2273866	102.354750	128.782445	KIAA1094	0.38	0.00	0.01	0.06	0.958	0.822	0.359
hCV3180154	rs2296793	103.236092	129.664612	DYT1	0.25	**1.05**	0.01	**1.81**	0.909	0.300	0.619
hCV2605168	rs1056171	104.299288	130.790555	ABL1	0.50	0.04	0.00	0.12	0.074	0.208	0.141
hCV8782344	rs1056899	105.681837	132.169455	KIAA0625	0.29	0.32	0.21	0.05	0.180	0.476	0.455
hCV32127084	rs7466085	106.130138	132.617936	N/A	0.49	0.55	0.14	0.00	0.893	0.933	0.826
hCV28004010	rs4962076	106.146282	132.634162	C9orf98/FLJ32704	0.20	0.36	0.01	0.46	0.222	0.996	0.560
hCV2980152	rs11243900	106.165871	132.653751	C9orf98/FLJ32704	0.41	0.55	0.05	0.68	0.784	0.651	0.699
hCV32127137	N/A	106.188696	132.676575	C9orf98/FLJ32704	0.17	0.41	0.51	0.00	0.136	0.127	0.212
hCV2535170	rs215156	106.205198	132.693075	C9orf98/hCG2033140	0.36	0.72	**1.00**	0.01	0.756	0.411	0.389
hCV1435374	rs4962218	106.264384	132.737397	C9orf98/FLJ32704	0.43	0.84	0.54	0.14	0.537	0.833	0.809
hCV2567972	rs214636	106.277080	132.764897	C9orf98/FLJ32704	0.43	**1.26**	0.31	0.41	0.617	0.942	0.910
hCV2536665	N/A	106.317774	132.805587	TSC1	0.46	0.93	0.55	0.36	0.272	0.988	0.591

hCV1247472	rs12551192	106.361322	132.849135	TSC1	0.24	0.08	0.06	0.23	0.825	0.937	0.974
hCV2253563	rs633153	106.404371	132.892186	GFI1B	0.45	**1.44**	**1.39**	0.01	0.913	0.651	0.996
hCV7582593	rs944204	106.416022	132.903834	GFI1B	0.46	0.26	0.08	0.05	0.130	**0.021**	**0.042**
hCV2535358	rs623489	106.445727	132.933551	GTF3C5	0.38	0.06	0.10	0.08	1.000	0.765	0.796
hCV2279860	rs685959	106.496869	132.984725	CEL	0.44	0.35	0.32	0.22	0.392	0.930	0.555
hCV2535450	rs886017	106.515825	133.003654	RALGDS	0.50	0.80	0.08	0.63	0.652	0.527	0.720
hCV2535940	rs671050	106.543639	133.031520	RALGDS	0.29	0.10	0.03	0.04	0.064	0.206	0.152
hCV2535973	rs2073927	106.573021	133.060900	GBGT1/RALGDS	0.47	**1.17**	**1.34**	0.11	0.808	0.228	0.090
D9S2157		106.577163			11 alleles	**2.86**	**2.16**	0.28	N/A	N/A	N/A
hCV2980279	rs9411461	106.582405	133.070281	FS/GBGT1	0.41	**1.98**	**1.19**	0.04	0.279	0.769	0.491
hCV2980256	rs10901243	106.619735	133.107929	OBP2B	0.36	**2.14**	**1.40**	0.00	0.188	0.695	0.081
hCV27224742	N/A	106.630889	133.119083	OBP2B	0.22	**1.50**	0.20	0.18	0.609	**0.026**	0.756
hCV3183098	rs2073824	106.674547	133.162187	ABO	0.34	0.43	0.69	0.00	0.940	0.907	0.600
hCV3183164	rs529565	106.691406	133.179054	ABO	0.36	0.32	0.24	0.00	0.619	0.198	0.842
hCV3183233	rs120858	106.751315	133.230364	SURF5/SURF6	0.49	**1.87**	0.80	0.00	0.542	0.705	0.610
hCV3183190	rs1179037	106.788988	133.268063	SURF4	0.50	**1.49**	0.97	0.03	0.891	0.442	0.967
hCV8784811	rs943623	106.817940	133.296925	XPMC2H	0.48	**1.32**	0.90	0.08	0.874	0.312	0.860
hCV11572323	rs3118663	106.832326	133.311307	XPMC2H/ADAMTS13	0.50	**1.53**	0.58	0.16	0.680	0.391	0.665
hCV3183371	rs652600	106.861602	133.340571	ADAMTS13	0.29	**1.02**	0.50	0.18	0.929	0.801	0.862
hCV8784809	rs1055432	106.874832	133.353793	ADAMTS13	0.34	0.31	0.32	0.03	0.374	0.795	0.540
hCV2536686	N/A	106.955316	133.436844	ADAMTSL2	0.24	0.17	0.00	0.30	0.745	0.793	0.248
hCV1247496	rs1105633	106.974470	133.456326	ADAMTSL2	0.44	**1.80**	0.99	0.29	0.244	0.088	0.538
hCV2971472	rs1029372	107.025317	133.507996	N/A	0.48	0.77	0.27	0.77	0.667	0.714	0.741
hCV2535803	rs2519148	107.041440	133.524120	DBH	0.47	**2.18**	**1.47**	0.04	**0.029**	**0.006**	0.079
hCV2535694	rs1611122	107.055817	133.538486	DBH	0.47	**2.73**	**1.36**	**1.05**	0.776	0.786	0.965
hCV2535675	rs2073837	107.069808	133.552482	DBH/SARDH	0.30	0.71	0.70	0.00	0.869	0.795	0.985
hCV11572672	rs1076149	107.095531	133.579210	SARDH	0.36	0.69	0.78	0.00	0.799	0.756	0.760
hCV2540688	rs495464	107.120467	133.603202	SARDH	0.43	0.17	0.20	0.00	0.788	0.923	0.925
hCV2537353	rs916620	107.143594	133.626304	SARDH	0.34	**1.59**	0.47	0.60	0.199	**0.029**	0.254
hCV12020823	rs1980852	107.936116	134.647976	hCG1814720	0.42	**1.12**	0.09	**1.41**	0.608	0.787	0.422
hCV3241385	rs968569	108.910258	135.619073	MGC29761/MRPS2	0.49	0.92	0.88	0.25	0.893	0.596	0.374
hCV12019285	rs6563	109.903628	136.665021	NOTCH1	0.47	0.40	0.54	0.22	0.295	0.656	0.105
hCV469299	rs11137268	110.905166	137.656213	FLJ31318	0.42	0.10	0.03	0.07	0.528	0.897	0.815
hCV247127	rs2229948	111.537898	138.284713	CACNA1B	0.26	0.00	0.00	0.01	0.255	0.080	0.532

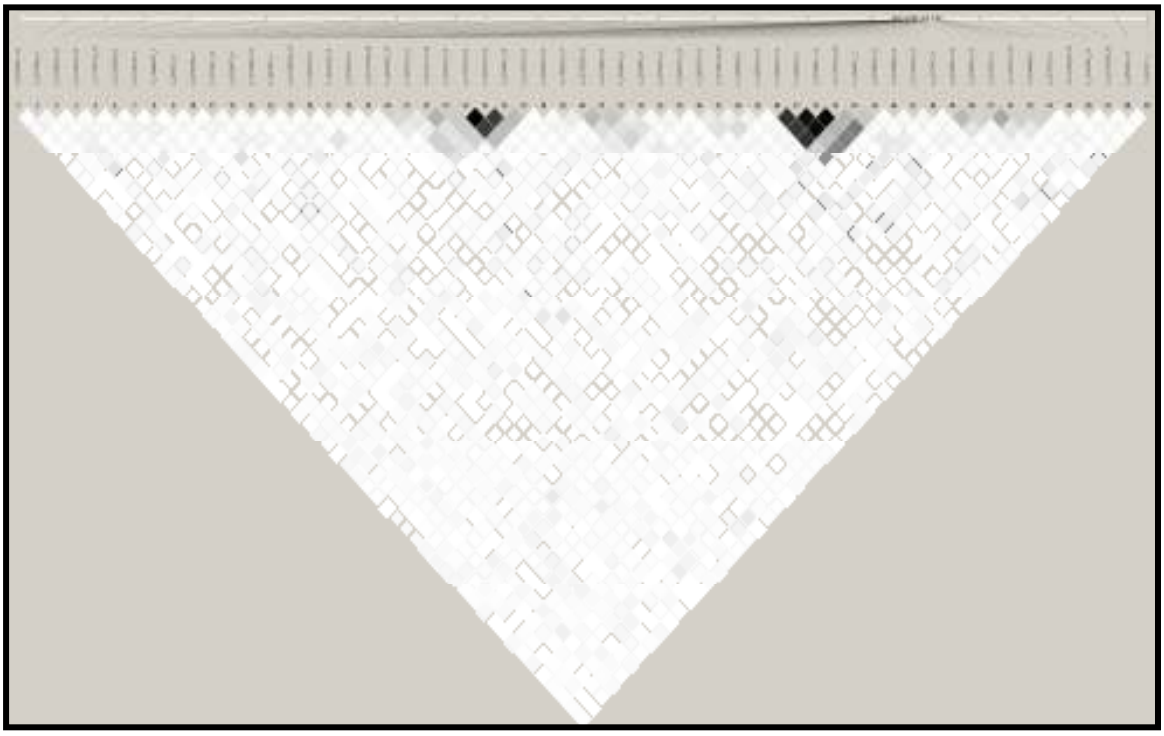


Figure 1. Chromosome 9q34 Haploview LD Plot – Before SNP Elimination The plot is shown before SNPs were removed to eliminate LD between markers to be used in multipoint analyses.

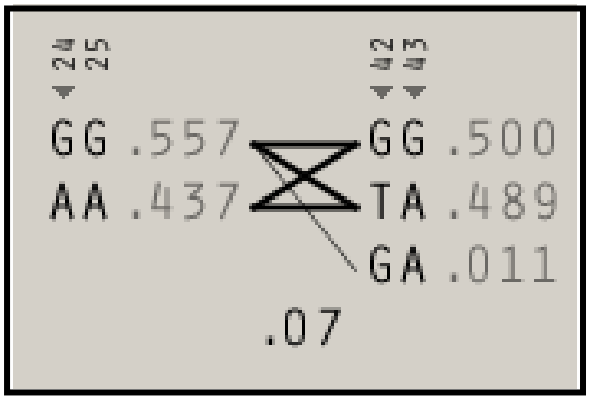


Figure 2. Chromosome 9q34 Haploview Haplotype Plot Two haplotype blocks with a multilocus D' of 0.07 were identified in the MS data set. Population frequencies are shown to the right of each haplotype and htSNPs are indicated by triangular pointers.

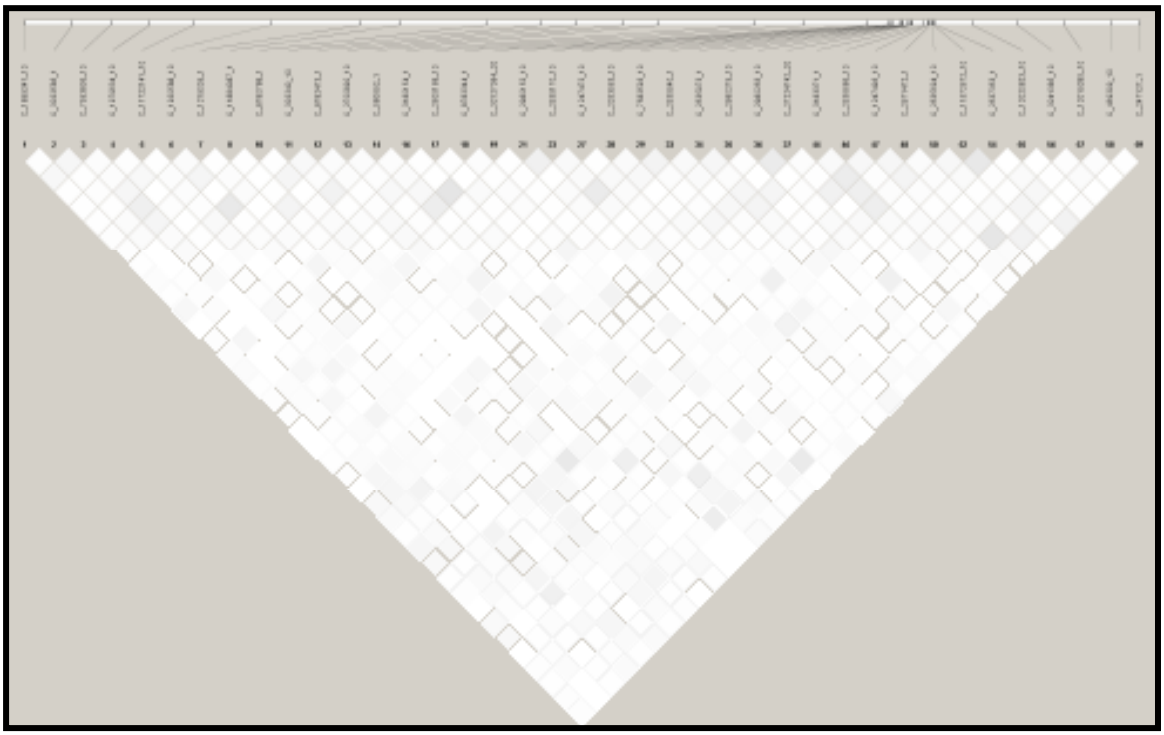


Figure 3. Chromosome 9q34 Haploview LD Plot – After SNP Elimination The plot is shown after SNPs were removed to eliminate linkage disequilibrium between markers to be used in multipoint analyses.

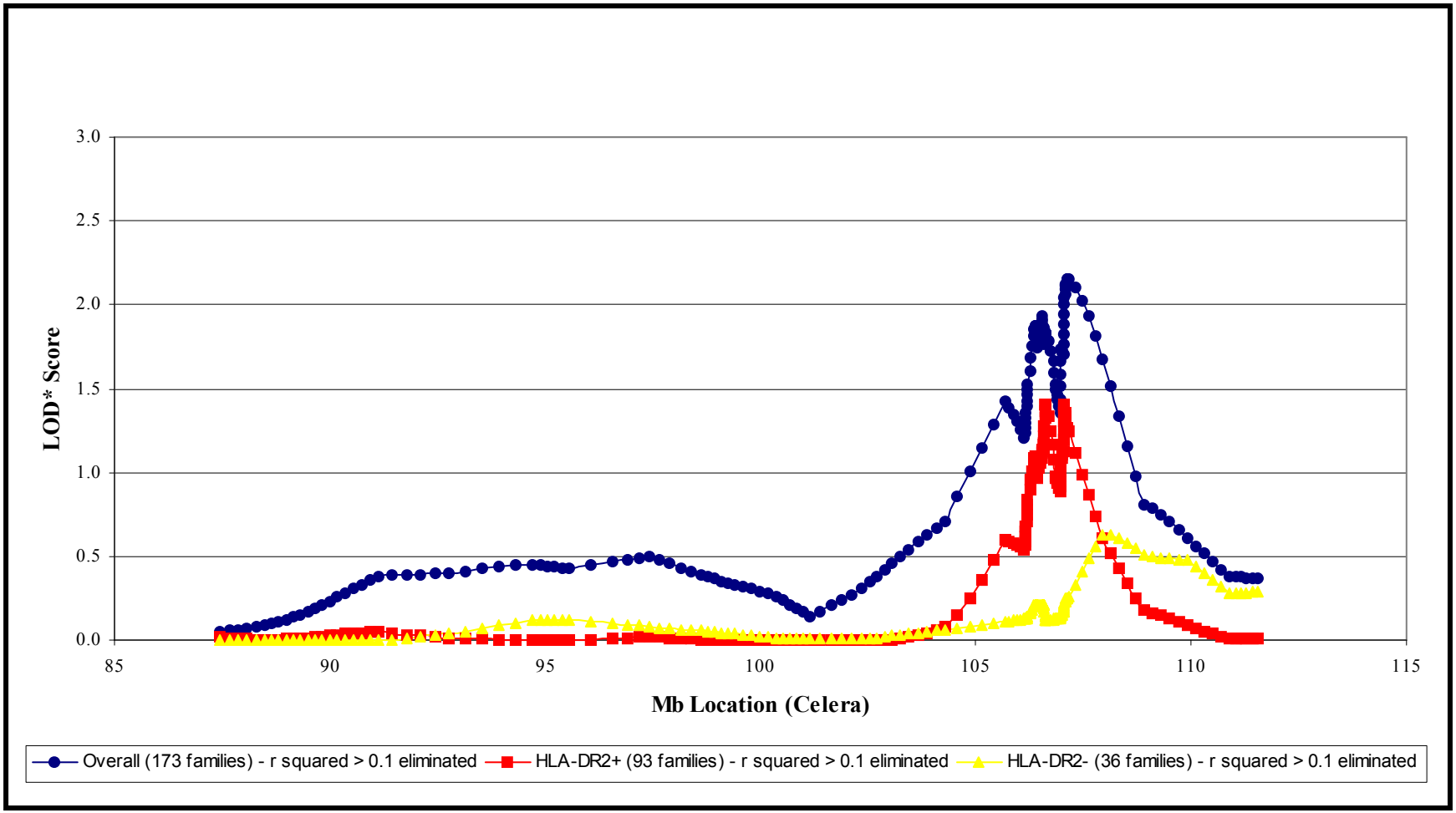


Figure 4. Chromosome 9q Multipoint Linkage Plot

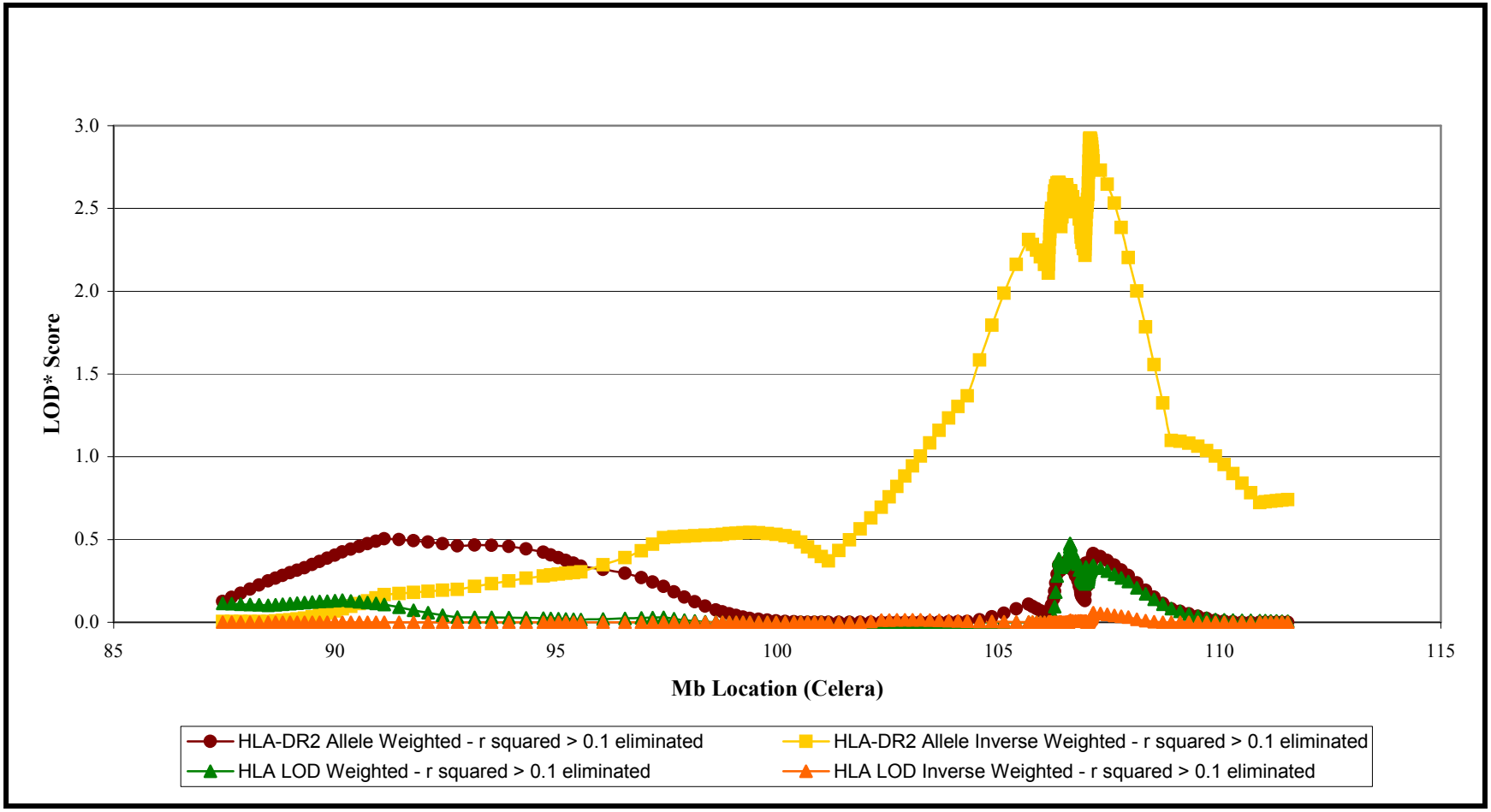


Figure 5. Chromosome 9q Multipoint Linkage Plot – HLA Conditional Analysis

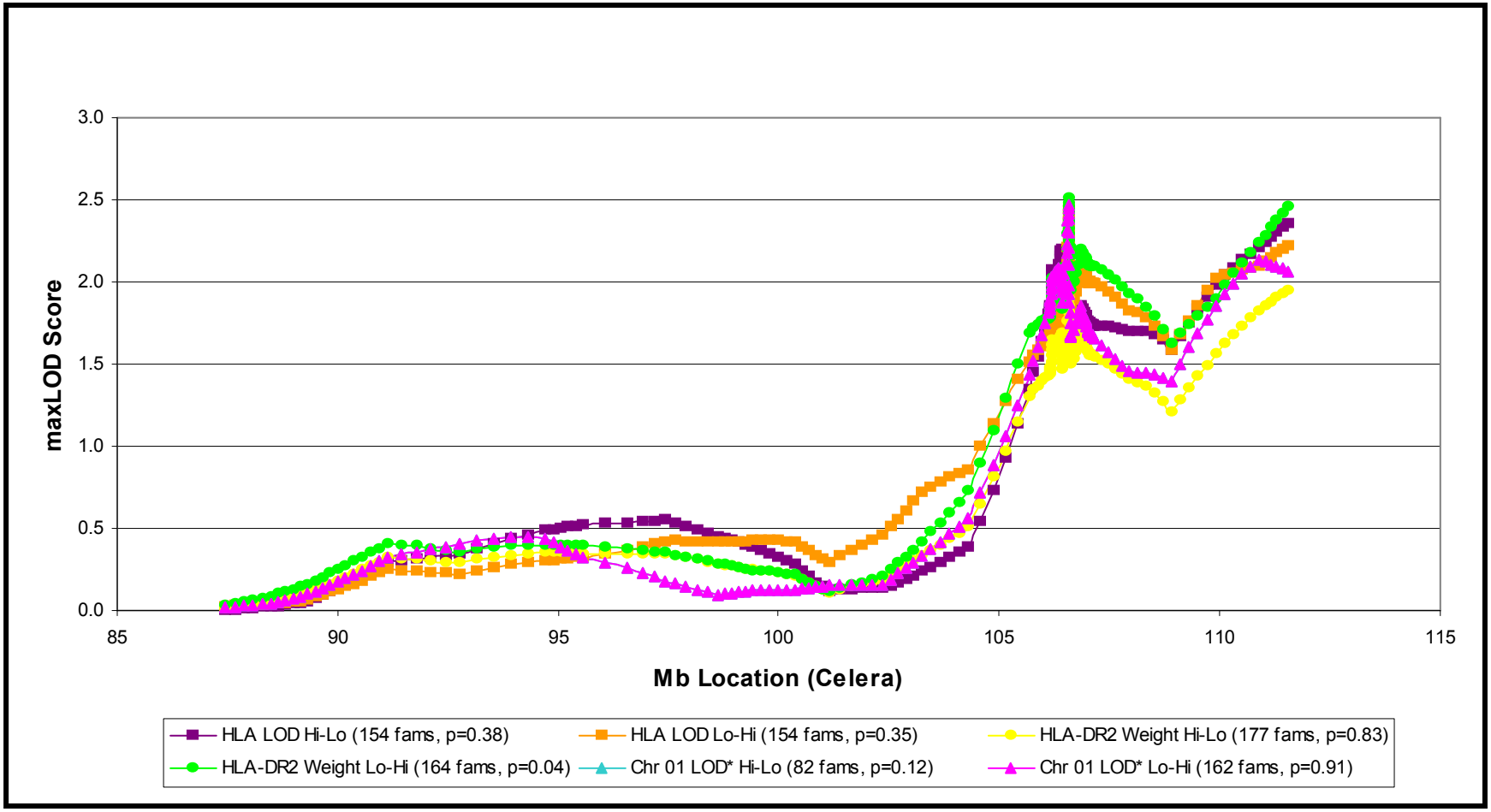


Figure 6. Chromosome 9q Multipoint Linkage Plot – Ordered Subset Analysis

CHAPTER VII

A DIRECTED INVESTIGATION OF CHROMOSOME 1Q44[**]

Abstract

Genomic linkage screens have served as the workhorse of genetic studies for complex diseases over the past decade. Despite the success of these screens in identifying regions of interest for MS, the thorough follow-up of candidate genes in all regions of linkage from a typical screen is prohibitive in terms of both time and cost—and has ultimately failed to identify MS susceptibility genes outside of the MHC. In addition to the prohibitive scale of these studies, it is likely that the candidate gene approach has been hampered by the incomplete identification and characterization of genes or regulatory elements that are directly related to disease pathophysiology. In this study, we employed a novel approach to investigate genetic association on the scale of a linkage peak for MS. We hypothesized that focusing on SNPs located in evolutionarily conserved regions would increase the likelihood of detecting variants that are associated with MS. This approach entailed the identification of multi-species conserved sequences and the development of a system to prioritize SNP selection. Use of this new approach on chromosome 1q44 resulted in the identification of four subregions demonstrating significant association with MS susceptibility.

[**] Chapter adapted from:

R03 Grant Submission: A Conserved Sequence Approach for MS Association Studies (Dr. Douglas P. Mortlock and Shannon J. Kenealy)

Introduction

Chromosome 1q44 was selected for investigation in this study based on several lines of evidence. Marker D1S547 in this region met the criterion for further interest in both two-point and multipoint calculations in the MSGG genomic screen (HLOD = 2.17; Z score = 3.38) [(105)] (see Chapter 4). Follow-up studies continued to demonstrate evidence of linkage in this region and narrowed the interval of interest to ~ 7.0 Mb (corresponding to a – 2.0 LOD score confidence interval) (Figure 1) (see Chapter 5). Markers in the 1q44 region have also demonstrated suggestive linkage and/or association in several other genomic screens conducted in a variety of MS study populations [(44); (47); (52); (53); (94)]. Another compelling piece of evidence is linkage to this region in studies for the autoimmune diseases rheumatoid arthritis and systemic lupus erythematosus, suggesting the presence of a gene involved in general autoimmune processes [(95); (110); (111); (112); (113); (96); (114)]. Interestingly, all of the markers demonstrating linkage in the MS, rheumatoid arthritis, and systemic lupus erythematosus screens are within a few Mb of the peak markers from the MSGG genomic screen and follow-up study.

As discussed in Chapter 5, follow-up of candidate regions identified in the MSGG genomic screen entailed genotyping SNPs at ~ 1 Mb intervals flanking ≥ 10 Mb on each side of peak screen markers in an expanded U.S. data set (Kenealy et al., manuscript submitted). Analyses in the 1q44 region in the follow-up study not only continued to demonstrate evidence for linkage (LOD = 2.99 in multipoint calculations for the overall data set), but also demonstrated a substantially narrowed linkage interval [(122); (123); Kenealy et al., manuscript submitted]. This narrowed interval is ~ 3.5 Mb for a LOD

score cut-off of 3.5 (corresponding to a – 1.0 LOD score confidence interval) and ~ 7.0 Mb for a LOD score cut-off of 2.5 (corresponding to a – 2.0 LOD score confidence interval) (Figure 1).

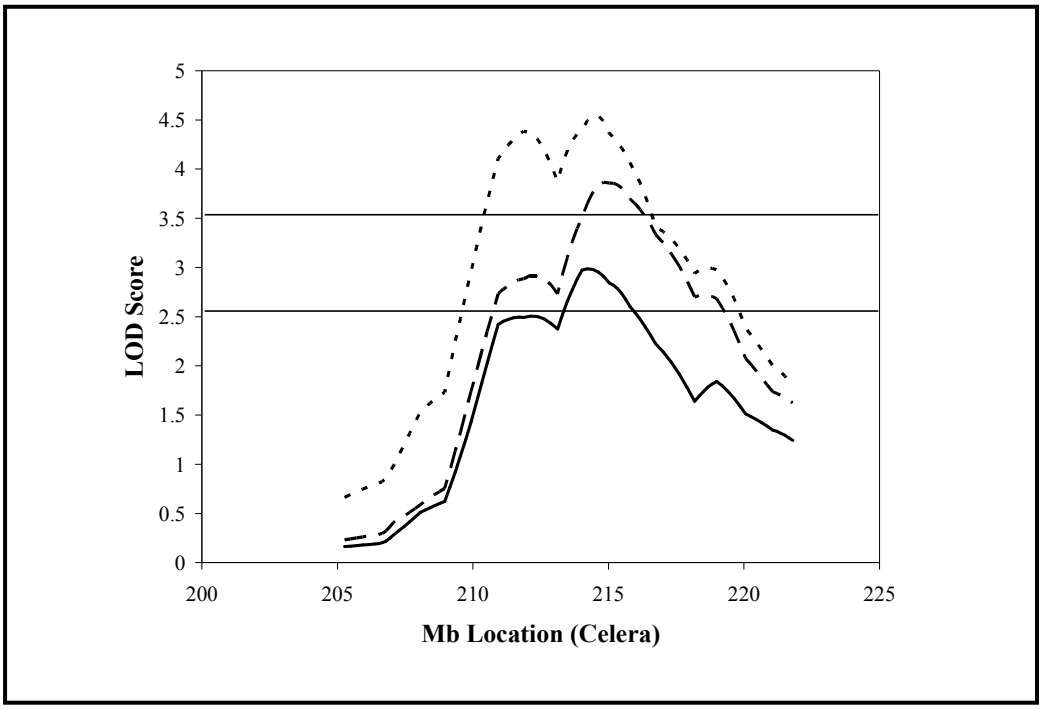


Figure 1. Multipoint Linkage Analyses for Chromosome 1q Original multipoint calculations are represented by the solid curve. Multipoint calculations for an OSA subset identified by ranking families for HLA effects are represented by the dashed curve. Multipoint calculations for an OSA subset identified by ranking families for chromosome 2 effects are represented by the dotted curve. Cut-offs corresponding to – 1.0 and – 2.0 LOD score confidence intervals are indicated by the two horizontal lines.

Despite the success of linkage analysis in substantially narrowing the region of interest on chromosome 1q44, the power of linkage studies to detect moderate genetic effects in small chromosomal regions is limited. This study therefore focused on finer mapping of the 1q44 region with allelic association methods testing for linkage disequilibrium (LD).

The number of assays necessary to thoroughly investigate association in genomic screen regions can be prohibitively large. Even in the narrowed region on 1q44, large-scale genotyping is required to conduct a thorough association study. Traditional approaches have attempted to reduce the amount of genotyping in these studies by testing only markers located in candidate genes. However, even the narrowed interval of ~ 3.5 Mb on 1q44 contains 28 known or predicted genes reported in the Celera and public databases, several of which could serve as candidate genes based on their proposed function (e.g. involvement in autoimmunity, viral susceptibility, oxidative stress/mitochondrial function, or neuronal processes) or based on their tissue expression patterns (e.g. in brain and spinal cord). A brief description of a few of these candidate genes is provided in Table 1. In addition, many of the remaining genes in this region have undergone little or no functional characterization, which would likely result in failure to select these genes for investigation in candidate gene studies despite their potential involvement in disease.

In this study, we employed a novel approach that incorporates evidence from positional and functional approaches to expedite follow-up studies in candidate regions. The utility of combining positional and functional approaches is evident in several recent studies for complex genetic disorders. The emerging concept of genomic convergence suggests that parallel investigations of genetic linkage, association, and expression data will speed disease gene discovery [(124)]. Recent application of this process to prioritize candidate genes on chromosome 10 in Alzheimer disease and Parkinson disease led to the successful identification of two genes significantly associated with these diseases [(125)].

Table 1. Functional Candidate Genes in the ~ 3.5 Mb Region on Chromosome 1q44

Gene Name / Protein Name / Celera ID	Gene Symbol	Tissue Expression	Comments
hCG2038857	N/A	macaque brain	homology to neural cadherin-like cell adhesion genes
cholinergic receptor, muscarinic 3	CHRM3	human brain	role in neurogenesis
formin 2-like protein	FMN2	human brain, spinal cord	N/A
gremlin 2 / protein related to DAC and cerberus	PDRC	human brain	mapped to neurons and spinal cord in mouse; role in TGFβ signaling
regulator of G-protein signaling 7	RGS7	human brain	upregulated in spinal cord injury; role in neuronal excitability
fumarate hydratase	FH	N/A	mitochondrial precursor
kynurenine 3-monooxygenase	KMO	N/A	role in oxidoreductase activity
opsin 3 (encephalopsin, panopsin)	OPN3	human brain, spinal cord, thymus	N/A
choroideremia-like / Rab escort protein 2	CHML	human thymus	role in neurophysiology, regulation of balance
hCG2042651	N/A	N/A	homology to dopa decarboxylase
exonuclease 1	EXO1	human thymus	role in DNA binding/repair, response to endogenous stimulus
beclin 1-like protein	N/A	N/A	role in antiviral host defense
hCG2041433	N/A	N/A	homology to FRG1 gene (expressed in brain, lymphocytes)

Our study utilized conserved sequence approaches to identify potentially functional sequences on chromosome 1q44. These conserved sequence approaches have several potential advantages over more traditional approaches that focus solely on SNPs located in coding regions. For example, since noncoding regulatory elements that control neighboring genes can be dispersed across large areas devoid of coding sequences, conserved elements may help discriminate functional regions within large noncoding areas that do not share LD with coding markers [(126); (127); (128)]. Conservation can also indicate coding regions that lack strong annotation support, such as alternatively spliced exons, RNA genes, "novel" genes with no homology to other gene families, or genes expressed at very low levels such that transcriptional evidence is lacking. Finally, by using conservation to prioritize SNPs, the odds may be increased that "functional" SNPs impacting the phenotype in question will actually be genotyped. For example, variation in transcriptional levels for key genes might play a significant role in disease risk. SNPs in noncoding cis-regulatory sequences (e.g. enhancers, repressors, or chromatin structural regulators) might contribute to the genetic component of this process by modulating transcriptional output.

There are currently several publicly available tools to detect evolutionarily conserved sequences across large genomic regions by performing sequence alignments [(129); (130)]. However, simple pairwise sequence comparisons have drawbacks for use as a systematic approach in the prioritization of conserved regions. For example, in a relatively large region, sequence alignment between any two mammalian species can provide too much aligning sequence, resulting in the identification of large amounts of

sequences that are not preserved as a result of selective processes, and are thus not likely to be functional [(131)]. Conversely, sequence alignment between more divergent species (e.g. between human and fugu) can provide too little information, resulting in the identification of only highly conserved protein-coding regions, while virtually all noncoding regions fail to be detected.

To minimize these drawbacks, new methods of alignment that compare sequences from multiple species have been developed. These new multi-species conservation methods have significant promise for detecting functionally conserved genomic sequences. A recent study showed that human sequences that are likely to be functional can be detected with improved sensitivity and specificity by comparing sequence from three or more vertebrate species [(132)]. These new methods may therefore have the ability to greatly increase the signal-to-noise ratio for detecting conserved sequences.

An algorithm for detecting multi-species conserved sequences (MCS) was recently optimized for scoring multi-species alignment data across large genomic regions [(133)]. This algorithm allows MCS scores to be assigned across any human genomic region that has been aligned with sequences from multiple species. A major advantage of MCS analysis is that it assigns a score to every 50-base pair region of human sequence that can be aligned to any of the comparison species. Regions that show similarity across many species or have stronger percent identity in pairwise matches receive relatively higher scores.

Our collaborator, Dr. Elliott Margulies, has performed MCS analysis across the entire human genome and made this analysis available online through the National Human Genome Research Institute (NHGRI) website and the University of California Santa Cruz (UCSC) genome browser (see NHGRI WebMCS website; UCSC website). This analysis incorporates mouse, rat and chick genome sequence alignment data to assign MCS scores to 50-base pair windows across the human genome. Alignment results allow for prioritization of all sequences that fall above a defined threshold in a region of interest (e.g. the top 5% scoring 50-base pair sequences from a defined region) (Figures 2 and 3).

In this study, we formulated a systematic approach to expedite the follow-up of positional candidate regions identified through linkage studies. This approach incorporated MCS analysis (as a tool to prioritize selection of SNP markers), current high-throughput genotyping techniques, and powerful statistical analysis methods. The application of this novel genomic convergence approach to a linked region on chromosome 1q44 resulted in the identification of four subregions demonstrating evidence of association with susceptibility to MS.

Material and Methods

Families

The data set used in this detailed investigation of chromosome 1q44 focused on the 91 families demonstrating evidence of linkage to this region in the studies presented in Chapter 5 [(105); Kenealy et al., manuscript submitted]. The data set included individuals from the 91 linked families that were most informative for analyses of allelic

association. In addition, at least one trio or discordant sib-pair (DSP) from each of the unlinked families from the studies presented in Chapter 5 was included in the study. A total of 1012 MS samples and 44 controls (2 CEPH controls and 2 blinded interplate controls per 96-well plate) were genotyped.

Molecular Analysis

SNPs were selected as markers for this study because of their evolutionary stability (and therefore likelihood to preserve LD information), abundant location throughout the genome, and amenability to high-throughput genotyping. In addition, SNPs are thought to be the source of many risk variants for complex genetic diseases.

The Illumina BeadArray™ platform was selected for rapid and accurate SNP genotyping of samples. This platform was comparatively high-throughput and cost-effective for genotyping of our large data set. The Illumina system also provided automated outputs that eased transfer between data generation and the PEDIGENE® database system used in the statistical analyses. Assays were synthesized by Illumina and samples were individually genotyped by our collaborators at the Duke Genomics Resource Laboratory Core.

The web-based WebMCS tool was used to identify multi-species conserved sequences in the ~ 7.0 Mb region through alignment of mouse, rat, and chick sequence to human chromosome 1q44 sequence ([(133)]; see NHGRI WebMCS website; UCSC website).

SNPS located in 5% MCS sequences were identified through the overlap of the 5% MCS output and SNP output from the UCSC website (see UCSC website).

An Illumina custom SNP panel was designed to genotype 768 SNPs (consisting of two oligo pool assays of 384 SNPs). Selection of SNPs in conserved regions was based on informativeness (with preference given to SNPs with high minor allele frequencies), validation (with preference given to SNPs confirmed by multiple lines of evidence), location (with preference given to SNPs located at regular intervals), putative function (with preference given to SNPs in coding, splice site, and mRNA UTR regions), and scores provided by Illumina (with preference given to SNPs generating scores > 0.60). Illumina scores were determined by an algorithm weighing a series of factors to predict the success of each locus within an oligo pool assay. Scores ranged between 0 and 1, with Illumina recommending selection of SNPs generating scores > 0.60.

Quality Control

All samples were quantified using a TaqMan® RNase P Detection kit and were concentrated or diluted to yield a final concentration of 50 ng/ul (Applied Biosystems, Foster City, CA). Genotyping of the 768 SNPs required < 500 ng total DNA per sample.

Duplicate quality controls samples were placed both within and across PCR plates and equivalent genotypes were required for all quality control samples to ensure accurate genotyping. Hardy-Weinberg calculations were performed for each marker and Mendelian inconsistencies were identified using PedCheck ([(134)]; [(90)]). Suspect samples and genotypes were dropped from the analyses. All SNPs were required to have

> 95% of possible genotypes. Verification of relationships between pairs of samples within families was performed using RELPAIR [(91)].

Statistical Analysis

SNP genotypes were analyzed for LD using Haploview software [(137)]. Data was analyzed with the pedigree disequilibrium test (PDT) for individual effects and Haploview for multilocus effects [(81)]. The PDT and Haploview are both powerful methods and are complimentary to each other. The PDT uses marker genotype data from DSPs or affected individuals and their parents (provided that both parents have been genotyped). The PDT can utilize information from related DSPs and/or parents and affected offspring from extended pedigrees. The tests provided by the PDT are independent of linkage, and can therefore extract additional information from the families used in linkage analysis—making this a robust method for single locus analyses [(135)]. Haploview was performed to simultaneously examine multiple SNPs and construct LD plots of the region. Measures of LD generated in Haploview are reported in terms of p values [(137)].

Statistical Power Calculations

Because very little is known about the underlying genetics of MS, it is difficult to determine the power of genetic analyses for the disease. With this caveat in mind, the MSGG performed simulations to test the power of the PDT to detect genetic effects. PDT power calculations were performed using an alpha value = 0.05 and a disease allele frequencies of 0.04 and 0.10 (corresponding to λ_s values = 1.25 and 1.75, respectively).

Assuming that the marker allele being tested is the disease susceptibility allele, the PDT was found to have good power to identify a locus with as few as 150 DSPs, even when the allele confers only a small genetic effect for the disease. Additional simulations were performed for the more likely scenario that the SNP allele being tested is in LD with the disease susceptibility allele. Not surprisingly, power was found to decrease with the degree of association. For a sibling recurrence risk (λ_s) of 1.25, the PDT provides reasonable power when association is only 60% of its maximal value for 150 DSPs. With a sample size of 400 DSPs (less than the number available in our data set), the PDT provides 80% power with as little as 40% of maximal association. In addition to DSPs, our data set also includes many trios. Studies have shown that trios are at least as powerful as DSPs under most conditions [(135)].

Because multilocus tests generally incorporate more information than single locus tests, we expected the power of Haploview to be similar or greater than the power of the PDT. In order to assess the power of multilocus association statistics, we performed calculations using the PBAT software package [(136)]. The PBAT program was used to determine power for a range of genetic models in our available data sets (Table 2). PBAT power calculations were performed using an alpha value = 0.05 and disease allele frequencies of 0.2 and 0.01 (to model common susceptibility alleles for autosomal dominant and autosomal recessive disease models, respectively). As with the PDT, family based association tests were found to have good power to identify a locus in any of our multiplex or simplex data sets when assuming the marker allele being tested is the susceptibility allele, even when the susceptibility allele confers only a small genetic

effect for the disease. When assessing the more likely scenario that the SNP allele being tested is in LD with the disease susceptibility allele, we again detected reasonable power in the multiplex data set when association is only 40% of its maximal value and marker allele frequencies are as low as 0.05 (under both disease models). Reasonable power was also seen for comparable scenarios in the simplex data sets, with the African-American simplex data set demonstrating slightly reduced power compared to the U.S and U.K simplex data sets. As expected, power was found to decrease with the degree of association and with the level of marker informativeness in each data set.

Table 2. Available Data Sets

Family Type	# Families	# Affecteds	# Unaffecteds
Multiplex	192	492	881
Simplex (U.S.)	593	593	1305
Simplex (U.K)	1000	1000	2000
Simplex (Af-Am)	489	489	339

Results

A preliminary assessment of the 1q44 region was performed using tools available through Celera and several public databases (see Celera website; UCSC website; Ensembl website; NCBI website). Database mining of the ~ 7.0 Mb region revealed 42 known or predicted genes and 24,977 SNPs (Table 1; Table 3).

A graphical display of the UCSC genome browser output, including tracks for chromosomal bands, known genes (based on information from the SWISS-PROT,

TrEMBL, mRNA, and RefSeq databases), Ensembl genes, GenBank mRNAs, and regions generating 5% MCS scores, is provided in Figure 2 (see UCSC website). An example of text output for 5% MCS regions in the UCSC genome browser is provided in Table 4. A detailed classification of SNPs located in the 1q44 region is provided in Table 3.

Previous analyses suggest that the top ~ 4-7% of MCS scores are very likely to indicate regions undergoing evolutionary selection [(133)]. This threshold also detects the vast majority of coding exons, while still detecting many noncoding regions. We therefore selected SNPs from the top 5% MCS-scoring regions from the ~ 7.0 Mb positional candidate region on chromosome 1q44. An overview of this output revealed 900 SNPs located in 5% MCS sequences from this region and indicated that in addition to detecting many coding exons, numerous 5% MCS regions are indeed within intronic and intergenic areas (Table 3).

Several SNPs located in 5% MCS regions were not selected for the study due to the nature of the variation (e.g. insertion/deletions or multiple mutation events leading to > 2 alleles) and/or the failure to generate an Illumina score > 0.60. Because elimination of these SNPs resulted in the identification of less than 768 SNPs in conserved regions, additional SNPs were selected from non-conserved regions with a similar prioritization scheme. Average spacing of the 768 SNPs in the ~ 7.0 Mb region on 1q44 region was < 10 kb—allowing for coverage of the region that is appropriate for observed patterns of LD in Caucasian populations.

Table 3. Classification of SNPs in the ~ 7.0 Mb Region on Chromosome 1q44. SNP classifications and heterozygosity information were obtained from dbSNP build 124 (see NCBI dbSNP website). Heterozygosity information was not available for all SNPs.

Location / Classification	# SNPs
~ 7.0 Mb Region	24977
5% MCS Sequences	900
Heterozygosity > 0.10	4548
Heterozygosity > 0.20	3799
Heterozygosity > 0.30	3051
Heterozygosity > 0.40	2152
Heterozygosity > 0.50	12
Coding Synonymous	39
Coding Nonsynonymous	72
Intronic	10133
Splice Site	2
mRNA UTR	1743

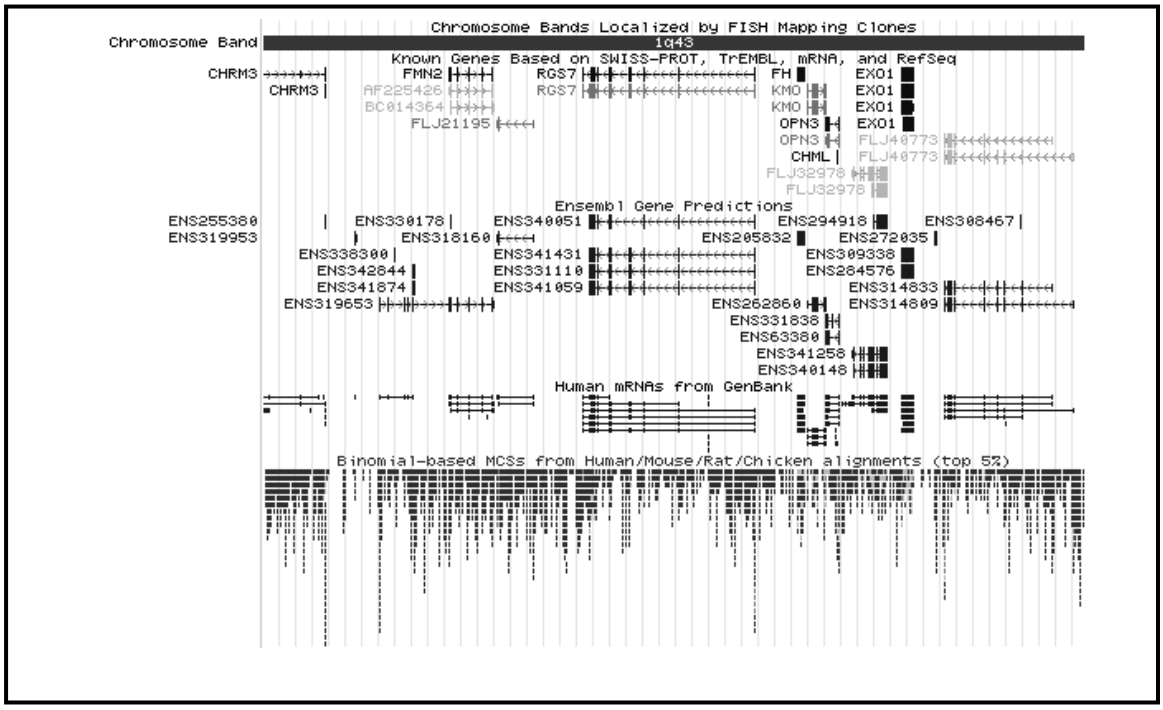


Figure 2. UCSC Genome Browser for the ~ 7.0 Mb Region on Chromosome 1q44

Table 4. Example of Coordinates for the 5% MCS Output in the UCSC Genome Browser Base pair positions are shown for a small fraction of conserved regions identified using the multi-species conserved sequence algorithm in the 1q44 region. The output for conserved sequence intervals was combined with SNP location information to identify markers to be tested for association with MS.

chrom	chromStart	chromEnd	name
chr1	236909187	236909248	mcs
chr1	236909280	236909322	mcs
chr1	236912128	236912326	mcs
chr1	236912334	236912361	mcs
chr1	236913252	236913283	mcs
chr1	236916559	236916585	mcs
chr1	236916710	236916743	mcs
chr1	236916769	236916805	mcs
chr1	236917352	236917387	mcs
chr1	236917448	236917483	mcs
chr1	236918129	236918161	mcs
chr1	236918164	236918206	mcs
chr1	236922118	236922160	mcs
chr1	236922217	236922243	mcs
chr1	236922475	236922504	mcs
chr1	236922509	236922535	mcs

Extensive quality control measures were taken to ensure consistent quality of DNA samples and SNPs. Of the 1012 MS samples genotyped in the study, twelve samples were eliminated from the analyses as a result of poor sample quality or Mendelian inconsistencies [(90)]. Of the 768 SNPs genotyped in the study, 189 SNPs were eliminated from the analyses. Seventeen of these SNPs failed to sufficiently amplify or cluster, while 172 SNPs were monomorphic (and therefore uninformative) in our MS data set. Of the remaining 574 SNPs, five SNPs demonstrated genotyping efficiencies between 90% and 95%. Of the 569 SNPs with genotyping efficiencies > 95%, only five SNPs fell below the predetermined HWE threshold of p values < 0.001. Given the apparent quality of genotyping (e.g. strong signals and tight clustering) for these SNPs, all five SNPs were included in the analyses. The average genotyping efficiency for the 569 analyzed SNPs was 99.6%. A complete list of SNPs passing all quality control measures is provided in Appendix D.

Haploview identified 124 LD blocks for the overall data set using the "solid spine" criteria (Figure 3). P values < 0.05 were generated for 39 LD blocks, with the most significant block generating a p value of 0.0011. Haploview identified 129 LD blocks for the subset of linked families only. P values < 0.05 were generated for 31 LD blocks, with the same block from the overall data set generating a p value of 0.0011. This block demonstrating association in the overall data set and linked subset contains 2 SNPs spanning a region containing the opsin 3 (OPN3) and opioid binding protein/cell adhesion molecule-like (OPCML) genes.

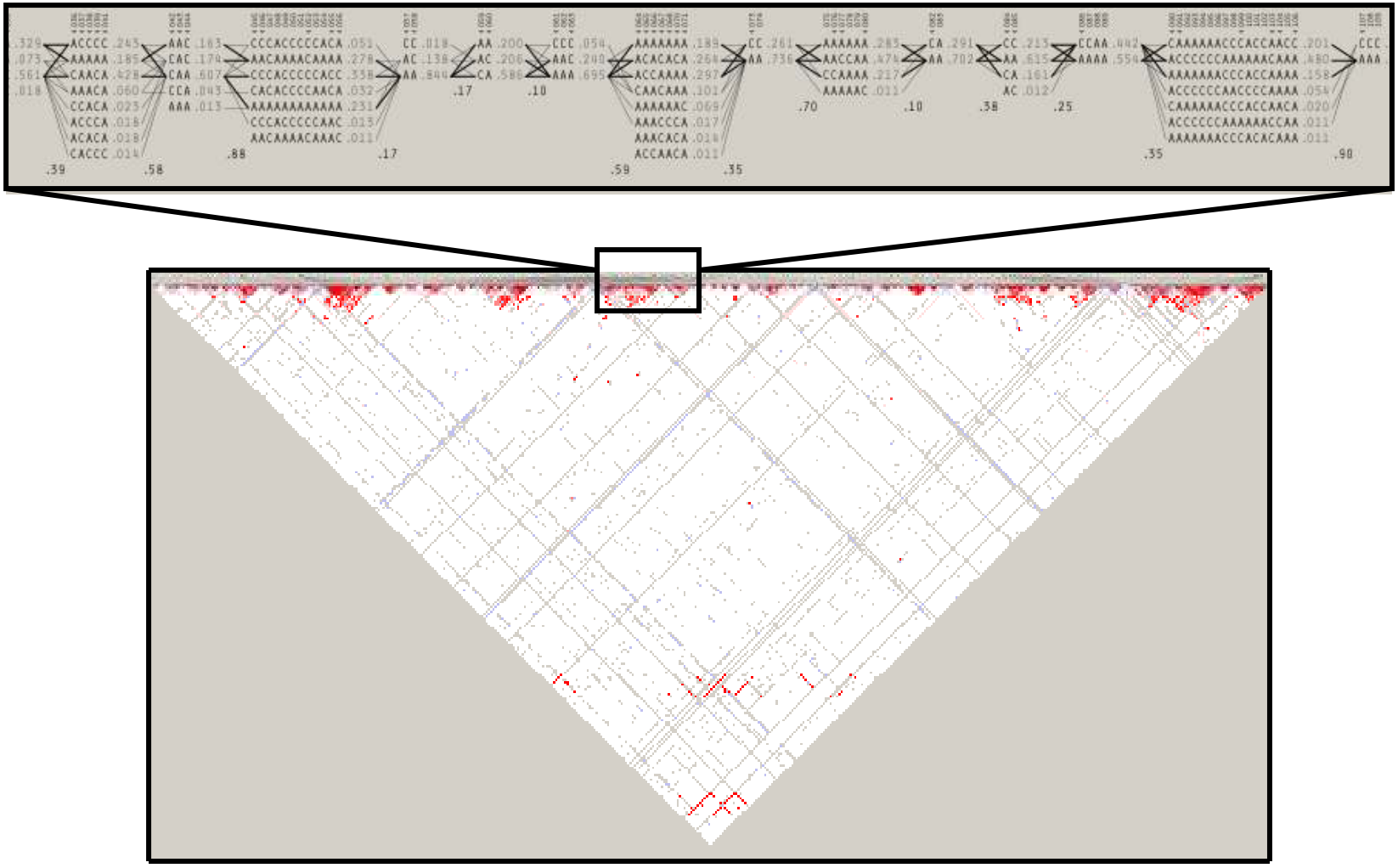


Figure 3. Haploview Plot and Representative Haploview Blocks for Chromosome 1q44

Single marker analyses performed with the TDT and PDT identified four subregions containing clusters of five or more SNPs demonstrating p values < 0.05 (Table 5). The first region contains five SNPs spanning ~ 43 Kb and includes the cardiac ryanodine receptor 2 (RYR2) gene involved in calcium signaling.

The second region contains fourteen SNPs spanning ~ 55 Kb and includes the formin 2 (FMN2) gene involved in cytoskeletal formation and a gremlin homology (PDRC) gene that functions as a BMP antagonist. In addition to the clusters of SNPs generating p values < 0.05, this second region also contains a haplotype block generating a p value < 0.05.

The third region contains twelve SNPs spanning ~ 163 K and includes the regulator of G protein signaling 7 (RGS7) gene. Significant p values in this region were generated only in the subset of linked families, providing consistent evidence of linkage and association to markers in this subregion in a subset of families.

The fourth region contains eight SNPs spanning ~ 48 K and includes a putative WD repeat domain 64 (WDR64) gene with homology to genes that coordinate protein complex assembly. Two of the SNPs in this region that generated p values < 0.05 are classified as coding non-synonymous SNPs (see NCBI dbSNP website).

Table 5. 1q44 Subregions with Interesting Single Marker Association Results

Region	Size	# SNPs	# 5% MCS SNPs	Gene(s) / Putative Gene(s)	Gene Function	Comments
1	~ 43 Kb	5	2	RYR2	calcium signaling	N/A
2	~ 55 Kb	14	11	FMN2 PDRC	cytoskeletal formation BMP antagonist	$p < 0.05$ for haplotype
3	~ 163 Kb	12	7	RGS7	G protein signaling	$p < 0.05$ only in linked families
4	~ 48 Kb	8	4	WDR64	protein complex assembly	$p < 0.05$ for 2 coding non-synonymous SNPs

Discussion

The approach employed in this study utilized a novel method for identification and prioritization of markers to be genotyped in a genetic association study of a positional region of interest. By combining locational and putative functional information, this method aimed to speed the process of identifying an MS susceptibility gene on chromosome 1q44.

The use of WebMCS in this study revealed several interesting observations. For example, a substantial portion of multi-species conserved sequences on chromosome 1q44 are located in noncoding regions. This observation highlights the importance of the MCS tool in detecting potentially functional sequences from relatively large genomic regions (e.g. several Mb) without arbitrary consideration of gene annotation. Preliminary results in our data set suggest that WebMCS could be a powerful for predicting functional information for genetic association studies.

We believe that this novel approach for follow-up of linkage studies increased the likelihood of successfully identifying a genetic factor in the 1q44 region. The discovery of a gene conferring susceptibility to MS in one of the subregions identified in this study would suggest that this approach could serve as a model for locational candidate studies in MS and other complex diseases.

Future Directions

SNPs demonstrating significant association results in this study (p values < 0.05) will be

selected for additional genotyping in ~ 600 U.S. simplex families (Table 2). Genotyping of the simplex families will be performed using the ABI TaqMan® 7900HT system in conjunction with TaqMan® Assays-on-Demand℠ (when available) or TaqMan® Assays-by-Design℠. The simplex families will be analyzed separately from the multiplex families to allow for replication of results in an independent data set.

SNPs continuing to demonstrate significant association results in the ~ 600 simplex families (p values < 0.05) will also be tested in the U.K. simplex and African-American simplex data sets. In addition to serving as an additional data set, the African-American samples will also provide a unique opportunity to assess ethnic-specific LD patterns in the 1q44 region.

Availability of these independent data sets translates to considerable power for detecting genetic effects and replicating positive results. The large size of these data sets also allows for conditional analyses that can potentially increase the power to detect genetic effects by identifying homogenous subsets. Identification of subsets can also reduce time and cost for detailed follow-up studies. Potential covariates for studies in these data sets include HLA-DR2 genotype, age at onset, progression, clinical subset, and family type. Empiric p values will be used to determine whether increased signals generated in subsets are statistically significant.

Several other phenomena of interest can also be investigated with the MCS tool. For example, the tool can be used to assess the distribution of conserved SNPs in exons

versus introns, as well as the distribution of SNPs in conserved versus non-conserved regions. In addition, we are interested in investigating the relationship between 5% MCS regions and SNPs demonstrating association. For example, preliminary analyses of the four subregions demonstrating association with MS indicate that 62% of SNPs generating p values < 0.05 are located in 5% MCS sequences. We plan to utilize the MCS tool to investigate these and other phenomena to maximally characterize the architecture of the chromosome 1q44 region.

CHAPTER VIII

CONCLUSIONS

Despite overwhelming evidence for a strong genetic component in MS, identification of genes conferring disease susceptibility has largely eluded researchers. With the exception of the MHC (containing HLA), traditional linkage analysis and candidate gene approaches have demonstrated insufficient power to identify genes or epigenetic factors that modulate MS disease risk. The failure of these studies highlights the need for new approaches and methodologies to identify the remaining genetic effect in MS.

Studies suggest that, like most common complex diseases, MS susceptibility is the result of multiple genes acting either independently or interactively in their contribution to overall risk. In addition to this complex etiology, clinical heterogeneity is likely to be a confounding factor in studies of MS. Fortunately, recent advances in bioinformatics, genotyping technologies, and statistical analysis methods are providing researchers with the tools necessary to address a variety of challenges involved in identifying genes for complex genetic diseases.

The goal of the work presented in this dissertation was to identify non-MHC loci that harbor MS susceptibility genes. Our studies entailed a new genomic convergence approach incorporating information gained from positional (linkage and association) and functional (comparative sequence) studies. In conjunction with high-throughput

genotyping and powerful new statistical analyses methods, this approach was used to conduct a directed investigation of the genetic contribution to MS. The following aims were undertaken in our genomic convergence approach to investigate the genetic susceptibility of MS:

Specific Aim 1: Conduct a simulation study using families generated with Genometric Analysis Simulation Program (GASP) software to assess the effectiveness of using haplotype-based positional mapping to define a minimum candidate region for a disease of interest.

Several variables (e.g. sample size, pattern of inheritance, and heterogeneity) were investigated for their effect on the power of the haplotype-based positional mapping approach. The modified consensus haplotyping approach in Specific Aim 1 demonstrated only modest power to narrow a minimum candidate region in the simulation data. Perhaps not surprisingly, the approach also demonstrated only a modest ability to narrow the minimum candidate region on chromosome 1q44 in MS genotyping data. Fortunately, more encouraging results for narrowing minimum candidate regions in the MS data set were simultaneously being generated in preliminary analyses of genotyping data for Specific Aim 2. Covariate analyses, including OSA, were therefore selected as the method to address genetic heterogeneity in studies of the MS data set in Specific Aim 2 (see Chapter 5).

Specific Aim 2: Test candidate genes for association with MS:

a. Identify and select a genetic interval of interest for MS.

We began our investigation with a genomic linkage screen that identified seven chromosomal regions of interest in a data set of multiplex MS families: 1q, 2q, 9q, 13q, 16q, 18p, and 19q. To narrow these regions, we developed an approach for more detailed linkage studies that capitalized on new methods for rapid and accurate genotyping of SNPs. In addition to increasing marker coverage in each region, we genotyped an expanded data set and devised covariate analyses schemes to account for genetic effect in the MHC (see Specific Aim 2b).

b. Apply the approach from Specific Aim 1 to the interval of interest identified in Specific Aim 2a.

As mentioned in the discussion of Specific Aim 1, covariate analyses were selected as the preferred method to address genetic heterogeneity in our data set. We developed subset, conditional, and ordered subset analyses schemes to account for genetic effect of HLA-DR and linkage to chromosome 1q. These follow-up analyses continued to provide evidence of linkage to several chromosomal regions, with particularly compelling evidence for chromosomal regions 1q44, 2q35, 9q34, and 18p11.

c. Identify candidate genes in the selected interval of interest.

Chromosome 1q44 was selected as the interval of interest for more detailed follow-up studies based on several lines of evidence, including continued evidence of linkage in Specific Aim 2b, evidence of linkage and/or association in other genomic screens conducted in a variety of MS study populations, and evidence of linkage to the region in studies of other autoimmune diseases. Although we identified candidate genes in the region based on their potential biological relevance to MS, we performed a detailed investigation of the 1q44 region using a novel approach detailed in Specific Aim 3.

d. Select a region and a narrow interval of interest for follow-up with SNPs in Specific Aim 3.

As mentioned in the discussion of Specific Aim 2c, chromosome 1q44 was selected for detailed follow-up studies in Specific Aim 3.

Specific Aim 3: Measure association between a dense population of SNPs and MS in the region of interest identified in Specific Aim 2.

a. Prioritize SNP markers for an MS association study in the interval of interest based on conservation between human, mouse, rat, and chick genome sequences.

We developed a systematic approach to expedite follow-up association studies for the positional candidate region on chromosome 1q44. In an attempt to increase the likelihood of detecting variants associated with MS, we developed and employed a novel method to identify and prioritize SNPs located in multi-species conserved sequences.

b. Measure association between a dense population of SNPs located in conserved regions of interest and a data set of families linked to the region.

Use of the method developed in Specific Aim 3a on chromosome 1q44 data resulted in the identification of four subregions demonstrating significant association with MS susceptibility. These regions contain several known or putative genes: ryanodine receptor 2 (RYR2), formin 2 (FMN2), regulator of G protein signaling 7 (RGS7), a putative WD repeat domain 64 (WDR64) gene, and a gremlin homology (PDRC) gene. Additional studies of the 1q44 region are currently underway.

The specific aims in this dissertation entailed a novel method for identification of genes underlying the susceptibility to MS. Use of these methods confirmed support for several chromosomal regions that warrant further investigation. It is our hope that the studies presented in this dissertation will result in the discovery of several genes associated with MS and that our genomic convergence approach will provide researchers with a method for unraveling the genetic heterogeneity of MS and other complex genetic diseases.

APPENDIX A

General Methods

Microsatellite Genotyping (Genomic Screen):

Primer Selection:

Obtain primer sequences for the desired microsatellite marker from the Genome Database (http://www.gdb.org/). If sequences are unavailable, design primers with Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Order synthesized custom oligos from Operon (http://www.operon.com/).

Primer Design:

Use Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) to design primers for the desired microsatellite marker:

> Obtain genomic sequence containing the desired microsatellite marker from the NCBI (http://www.ncbi.nih.gov/) or Celera (http://www.celeradiscoverysystem.com/index.cfm) website. Download FASTA-formatted sequence or add the symbol ">" to the beginning of the sequence (e.g. >ATCG).

Mask the FASTA-formatted sequence for repetitive DNA elements using RepMasker software (http://www.repeatmasker.org/).

Paste masked sequence into the Primer3 sequence window and specify targeted and/or excluded regions.

Change “Product Size Range” to include appropriate product size.
Change “Primer Size” to range from 18-24 bp (optimal: 21 bp).
Change “Primer Tm” to range from 55°C - 65°C (optimal: 60°C).
Change “Maximal Tm Difference” to 5°C.
Change “Primer GC%” to range from 40% - 60% (optimal: 50%).

Click on “Pick Primers” and retrieve forward and reverse primer sequences.

BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) the forward and reverse primer sequences to verify specificity of binding to the desired locus.

Helpful Hints: If Primer3 fails to return potential primer sets, ease constraints (e.g. widen ranges for “Primer Tm” and “Primer GC%”). If BLAST indicates binding to multiple loci (with high specificity), select a different primer set. When genotyping more than one microsatellite marker, consider designing primers for different product sizes to allow for multiplex reactions.

Polymerase Chain Reaction (PCR):

In each well of a 96-well PCR plate, combine 1.0 μL patient DNA (@ 0.02 μg/μl) with 9.0 μL PCR mix (shown below). Amplify DNA with the appropriate PCR program on the Hybaid© PCR Express Thermal Cycler or the MJ Research PTC-225® Peltier Thermal Cycler.

PCR Mix (96 rxns. @ 10 μL/rxn.):

880 μL GIBCO BRL® PCR SuperMix*

50 μL forward primer (@ 0.10 μg/μL)

50 μL reverse primer (@ 0.10 μg/μL)

*22 mM Tris-HCl (pH 8.4), 55 mM KCl, 1.65 mM $MgCl_2$, 220 μM dGTP, 220 μM dATP, 220 μM dTTP, 220 μM dCTP, 22 U recombinant *Taq* DNA Polymerase/ml, stabilizers

Annealing Temperature (AT) Calculation:

$(A + T) \times 2 = X$

$(G + C) \times 4 = Y$

$X + Y = Z$

Z – 5 = Annealing Temperature (AT)

PCR Program:

94°C – 4 min.

94°C – 15 sec.
AT – 30sec.
72°C – 45 sec.
} 35 cycles

72°C – 4 min.

4°C – ∞ (HOLD)

Gel Electrophoresis:

Size fractionate PCR products on a 6% denaturing polyacrylamide gel.

Glass Plate Preparation:

Spray one side of a long glass plate with TexClean™100. Remove TexClean™100 (and any adhered gel) with a razorblade. Spray the plate with ethanol and wipe clean with a KimWipe®.

Spray one side of a notched glass plate with ethanol and wipe clean with a KimWipe®.

Place the long plate on a gel slider with the clean side facing up. Place one plastic spacer on each side of the plate and secure the spacers at the top and middle of the plate with four plastic clips.

Place the notched plate on top of the long plate with the clean side facing down and overlapping the bottom of the long plate by one inch.

Gel Preparation:

6% Denaturing Polyacrylamide Gel:

75 mL National Diagnostics SequaGel® 6 Monomer Solution*
15 mL National Diagnostics SequaGel® 6 Complete Buffer** } ***

750 μL 10% Ammonium Persulfate

* 6 M urea, acrylamide, bis-acrylamide (19:1)

** 1X TBE (89 mM Tris Base, 89 mM Boric Acid, 2 mM EDTA, pH 8.3), TEMED

*** Combine SequaGel® 6 Monomer Solution and SequaGel® 6 Complete Buffer in the Monomer Solution bottle and invert several times. The combined solution can be stored in the Monomer Solution bottle for extended periods of time.

Stir the gel mix with the tip of a 60 cc plastic syringe (without creating bubbles). Draw mix into the syringe and start dispensing mix onto the edge of the notched plate. Continue to dispense mix while sliding the notched plate toward the top of the long plate. Remove clips when necessary and continue to dispense mix until the edges of the plates are flush.

Replace the four plastic clips on the sides of the plates. Insert plastic molding combs ½" into the space between the top of the plates to make a straight edge in the gel. Secure the molding combs with several metal clamps, with one metal clamp securing the position where the combs meet.

After the gel has polymerized (~ 20 min.), remove all clips and clamps from the plates.

Once a month (or as needed):

Spray the notched plate with a small amount of Acrylease™ Plate Coating. Buff Acrylease™ evenly over the notched plate with a KimWipe®. Remove excess Acrylease™ with ethanol and a KimWipe®.

Gel Rig Setup:

Place the plates in an electrophoresis rig. Secure the sides of the plates with plastic clips and the top of the plates with a plastic bar. Fill the top and bottom reservoirs with 1X TBE buffer. Place a loading spacer in the gap at the top of the gel and flush wells with an eyedropper.

Gel Loading:

Add 5 μL loading dye to each 10 μL PCR reaction. Denature PCR products for 3 min. @ 95°C. Load 2.5 μL of each sample on the 6% denaturing polyacrylamide gel with a 12-channel pipette. Run gel for ~1 hr. @ 75 W.

Loading Dye (10 mL):

9.5 mL formamide

500 μL 0.5 M EDTA

0.0012 g BromoPhenol Blue powder

SYBR® Gold Nucleic Acid Gel Stain (50 mL):

5 μL SYBR® Gold (10,000X concentrate in DMSO)

50 mL 1X TBE

Helpful Hint: Store SYBR® Gold dilution @ 4ºC for no longer than one week.

Separate the glass plates with a plastic plate splitter (the gel should remain on the long plate). Pour one 50 mL tube of diluted SYBR® Gold solution onto the gel. Let the SYBR® Gold solution sit on the gel for 5-10 minutes. Gently rinse the SYBR® Gold solution from the gel with H_2O and wipe the outside of the long glass plate dry. Scan the plate/gel on the Hitachi Biosystems FMBIOII laser scanner and print the image for genotyping.

SNP Genotyping (Genomic Screen Follow-Up):

Select SNPs from the Celera (http://www.celeradiscoverysystem.com/index.cfm) or Applied Biosystems (http://www.appliedbiosystems.com/) websites and order the corresponding Assays-On-Demand℠ or Assays-by-Design℠.

Assays-On-Demand℠:

PCR Mix:

1.0 μL patient DNA (@ 0.05 μg/μl)

2.5 μL TaqMan® Universal PCR Master Mix, No AmpErase® UNG (2X)

0.25 μL 20X TaqMan® SNP Genotyping Assay Mix

1.25 μL Milli-Q® H_2O

Helpful Hints: Protect fluorescent tags in the Assay Mix from light by wrapping each tube in aluminum foil. Minimize freeze-thaw cycles for the Assay Mix by simultaneously running multiple plates of DNA samples for each assay (or aliquot the assays in multiple tubes after the initial thaw).

Allelic Discrimination PCR Program:

95°C – 10 min.

92°C – 15 sec.
60°C – 1 min.
} 50 cycles

4°C – ∞ (HOLD)

Helpful Hint: SNP Genotyping Assay Mix can be diluted ≥ 4X for most assays (assess assay strength in the first PCR run and dilute mix for subsequent reactions accordingly).

Assays-by-Design℠:

PCR Mix:

1.0 μL patient DNA (@ 0.05 μg/μl)

2.5 μL TaqMan® Universal PCR Master Mix, No AmpErase® UNG (2X)

0.125 μL 40X TaqMan® SNP Genotyping Assay Mix

1.375 μL Milli-Q® H_2O

Helpful Hints: Keep fluorescent tags in the Assay Mix protected from light by wrapping each tube in aluminum foil. Minimize freeze-thaw cycles for the Assay Mix by simultaneously running multiple plates of DNA samples for each assay (or aliquot the assays in multiple tubes after the initial thaw).

Allelic Discrimination PCR Program:

95°C – 10 min.

92°C – 15 sec.
60°C – 1 min.
} 50 cycles

4°C – ∞ (HOLD)

Helpful Hint: SNP Genotyping Assay Mix can be diluted ≥ 4X for most assays (assess assay strength in the first PCR run and dilute mix for subsequent reactions accordingly).

Perform PCR amplification of Assays-On-Demand^{SM} or Assays-by-Design^{SM} in a 384-well GeneAmp® PCR System 9700 (Applied Biosystems) or a 384-well DNA Engine Tetrad® 2 Peltier Thermal Cycler (MJ Research).

Launch SDS 2.1 software on a computer connected to the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems). Designate the appropriate allele for each fluorescently tagged reporter dye (e.g. A for VIC, G for FAM). Place a PCR plate in the scanner and perform an endpoint read. Genotype samples using the 95% confidence interval function. Export genotyping data from the results table to a text document to be used in statistical analysis.

Other Reagents/Dilutions/Kits:

1X TBE Buffer (National Diagnostics):

89 mM Tris Base

89 mM Boric Acid (pH 8.3)

2 mM EDTA

Puregene™ DNA Hydration Solution:

10 mM Tris Base

1 mM EDTA

Milli-Q® H_2O (Millipore)

TaqMan® RNase P Detection Reagents Kit (Applied Biosystems)

General Methods of Collaborators:

HLA Genotyping (Genomic Screen):

Genotypes for HLA-DR in U.S. families were determined at UCSF using non-radioactive PCR-SSOP (Dynal Biotech).

Genotypes for HLA-DR in French families were determined using reverse dot blot hybridization.

APPENDIX B

Seven Candidate Regions for MS – SNP Information

Table 1. Chromosome 1q SNPs

Celera ID	Public ID	Celera Mb Location	Mb Interval	Gene	MAF	SNP Type	SNP
hCV1650709	rs734551	205.263268	1.114235	DISC1	0.48	Intronic	[A/G]
hCV2798667	N/A	206.377503	0.698733	MGC13186	0.50	Intronic	[A/G]
hCV8690690	rs701237	207.076236	0.985738	KCNK1	0.36	Intronic	[C/G]
hCV2711850	rs564212	208.061974	0.896874	hCG1813648	0.40	Intronic	[C/T]
hCV695532	rs423026	208.958848	0.971200	GNG4	0.49	Intronic	[C/G]
hCV7540624	rs1266380	209.930048	1.004790	LGALS8	0.47	Intronic	[A/G]
hCV16123987	rs2805432	210.934838	0.937254	RYR2	0.50	Intronic	[A/G]
hCV401377	rs1557132	211.872092	1.251360	hCG23440	0.50	Intronic	[T/C]
hCV2050524	rs2278644	213.123452	0.903277	hCG1984146	0.38	Intergenic/Unknown	[A/C]
hCV12008520	rs1934338	214.026729	1.004684	FLJ21195	0.47	Intronic	[C/G]
hCV605574	rs671989	215.031413	0.763360	CHML	0.48	Intronic	[A/G]
hCV26589316	N/A	215.794773	0.950948	hCG1989348	0.50	UTR3	[A/C]
hCV16244390	rs2490395	216.745721	1.438506	hCG16602	0.42	Intronic	[A/G]
hCV9487525	rs1566661	218.184227	0.807586	hCG1660865	0.45	Missense Mutation	[A/C]
hCV11669158	N/A	218.991813	1.067830	FLJ10157	0.41	Intronic	[T/G]
hCV15761414	rs3007406	220.059643	0.999964	FLJ23001	0.42	Intronic	[A/G]
hCV11669332	N/A	221.059607	0.737545	hCG1646765	0.48	Intronic	[C/T]
hCV12010310	rs2039824	221.797152		hCG1724360	0.36	Silent Mutation	[C/T]

D1S1594	214.036133 - 214.036253					Microsatellite	
D1S547	215.010921 - 215.011224					Microsatellite	
D1S1634	215.585528 - 215.585699					Microsatellite	

Table 2. Chromosome 2q SNPs

Celera ID	Public ID	Celera Mb Location	Mb Interval	Gene	MAF	SNP Type	SNP
hCV11827407	rs1982131	172.460711	1.103740	PDE11A	0.46	Intronic	[A/G]
hCV2152432	rs2847	173.564451	1.180400	SESTD1	0.43	Intergenic/Unknown	[C/T]
hCV31157131	rs13391691	174.744851	1.002856	N/A	N/A	Intergenic/Unknown	[G/T]
hCV1051244	rs720453	175.747707	0.964938	N/A	0.49	Intergenic/Unknown	[C/G]
hCV2074867	rs2128043	176.712645	0.885402	PDE1A	0.47	Intronic	[A/G]
hCV1543267	rs2952363	177.598047	1.050445	LOC129401/NUP35	0.45	Intronic	[G/T]
hCV426462	rs10497643	178.648492	1.112739	N/A	0.44	Intergenic/Unknown	[C/T]
hCV7617345	rs889909	179.761231	1.057397	N/A	0.33	Intergenic/Unknown	[G/T]
hCV11518651	rs10177153	180.818628	0.984987	N/A	0.36	Intergenic/Unknown	[C/T]
hCV1590313	rs1528233	181.803615	0.898701	CALCRL	0.42	Intergenic/Unknown	[A/G]
hCV474783	rs11692963	182.702316	0.772323	N/A	0.39	Intergenic/Unknown	[A/C]
hCV1729153	rs3134646	183.474639	1.140407	COL3A1	0.50	Intergenic/Unknown	[A/G]
hCV2831378	rs785260	184.615046	0.815149	N/A	0.49	Intergenic/Unknown	[A/C]
hCV22274194	rs3771300	185.430195	1.200225	STAT1;GLS	0.43	Intronic	[T/G]
hCV16207964	rs2356955	186.630420	0.405657	TMEFF2	0.45	Intronic	[A/C]
hCV541573	rs717621	187.036077	0.862366	N/A	0.49	Intergenic/Unknown	[A/T]
hCV11950973	rs726129	187.898443	0.936343	N/A	0.39	Intergenic/Unknown	[C/G]
hCV1346631	rs4591357	188.834786	0.871059	hCG2038662	N/A	Intronic	[A/G]
hCV8744355	rs1019845	189.705845	0.777310	Unknown	0.44	Intergenic/Unknown	[A/G]
hCV7620124	rs1551838	190.483155	1.129796	DNAH7	0.44	Intronic	[C/T]
hCV1291059	N/A	191.612951	1.093056	LOC91526	0.48	Intergenic/Unknown	[C/T]
hCV2123977	rs1064213	192.706007	1.187134	PLCL1	0.41	Missense Mutation	[A/G]
hCV2153217	rs2881208	193.893141	0.637732	hCG20092	0.35	Intergenic/Unknown	[C/T]
hCV3230597	rs1124639	194.530873	0.864581	FLJ38973	0.50	Intergenic/Unknown	[C/T]
hCV1223380	N/A	195.395454	1.122969	hCG1811467	0.46	Intergenic/Unknown	[C/T]
hCV347110	rs759419	196.518423	0.974936	ALS2CR7	0.48	Intergenic/Unknown	[A/G]
hCV8761981	rs3845802	197.493359	0.984214	LOC130026/WDR12/ALS2CR14	0.41	Intergenic/Unknown	[G/T]
hCV2821059	rs926169	198.477573	1.203750	CTLA4	0.41	Intronic	[G/T]
hCV1572836	N/A	199.681323	0.818677	ALS2CR19	0.48	Intergenic/Unknown	[C/T]

hCV2186447	rs1045043	200.500000	1.172078	NRP2	0.43	Intergenic/Unknown	[C/G]
hCV2772287	rs2287508	201.672078	0.731708	CPO	0.41	Intergenic/Unknown	[C/T]
hCV3090677	N/A	202.403786	0.880599	FZD5	0.43	Intergenic/Unknown	[A/C]
hCV1737148	N/A	203.284385	1.158995	hCG2041988	0.47	Intergenic/Unknown	[A/G]
hCV2668266	rs731953	204.443380	0.817609	hCG1821192/hCG2040260	0.50	Intergenic/Unknown	[G/T]
hCV2034024	rs2887914	205.260989	1.202672	CPS1/PRO0132	0.45	Intronic	[A/C]
hCV8835839	rs714393	206.463661	1.179599	ERBB4	0.50	Intronic	[C/T]
hCV1861868	rs1871946	207.643260	0.977608	ZNFN1A2	0.45	Intronic	[C/T]
hCV1552068	rs2372109	208.620868		hCG1648127	0.41	Intergenic/Unknown	[A/G]

GATA149B10	193.459832 - 193.459963			Microsatellite	
D2S1384	198.981939 - 198.982084			Microsatellite	
D2S1365	205.839346 - 205.839546			Microsatellite	

Table 3. Chromosome 9q SNPs

Celera ID	Public ID	Celera Mb Location	Mb Interval	Gene	MAF	SNP Type	SNP
hCV1899341	rs3810928	87.463072	1.030977	AMBP	0.47	Missense Mutation	[C/T]
hCV3022586	rs1061494	88.494049	0.820104	TNC	0.50	Missense Mutation	[C/T]
hCV7593836	N/A	89.314153	0.857951	EST-YD1	0.44	Intronic	[A/T]
hCV1979634	N/A	90.172104	0.945717	ASTN2	0.49	Intronic	[C/T]
hCV11722141	rs1927911	91.117821	1.650176	TLR4	0.31	Intronic	[A/G]
hCV1920588	rs1324623	92.767997	1.945431	DBCCR1	0.37	Intronic	[G/T]
hCV1219009	rs3747850	94.713428	0.849474	GSN	0.50	Intronic	[G/T]
hCV11884087	rs4679	95.562902	1.242257	NDUFA8	0.40	Silent Mutation	[T/C]
hCV58657	N/A	96.805159	0.625528	KIAA1608	0.39	Intronic	[C/T]
hCV8780788	rs1042486	97.430687	1.180163	LHX2	0.46	Silent Mutation	[C/G]
hCV302240	rs501963	98.610850	0.769140	RAB9P40	0.48	UTR5/Intronic	[A/G]
hCV8782473	rs2302748	99.379990	0.996620	PBX3	0.48	Intronic	[C/T]
hCV2700890	rs2874799	100.376610	0.778330	RALGPS1A	0.46	Intronic	[C/G]
hCV580692	rs514024	101.154940	1.199810	SH2D3C	0.45	Silent Mutation	[A/G]
hCV16180096	N/A	102.354750	0.881342	hCG30598	0.40	Missense Mutation	[A/G]
hCV3180154	rs2296793	103.236092	1.063196	DYT1	0.21	Silent Mutation	[A/G]
hCV2605168	rs1056171	104.299288	1.382549	ABL1	0.46	Silent Mutation	[A/G]
hCV8782344	N/A	105.681837	0.448301	KIAA0625	0.33	Missense Mutation	[C/T]
hCV32127084	rs7466085	106.130138	0.016144	N/A	0.46	Intergenic/Unknown	[A/G]
hCV28004010	rs4962076	106.146282	0.019589	C9orf98/FLJ32704	0.24	Intronic	[C/G]
hCV2980152	rs11243900	106.165871	0.022825	C9orf98/FLJ32704	0.39	Intronic	[A/G]
hCV32127137	N/A	106.188696	0.016502	C9orf98/FLJ32704	0.21	Intronic	N/A
hCV2535170	rs215156	106.205198	0.059186	hCG2033140/C9orf98	N/A	Intronic	[A/G]
hCV1435374	rs4962218	106.264384	0.012696	C9orf98/FLJ32704	0.49	Intronic	[A/G]
hCV2567972	rs214636	106.277080	0.040694	C9orf98/FLJ32704	0.49	Intronic	[A/G]
hCV2536665	rs2809245	106.317774	0.043548	TSC1	0.47	Intronic	[C/T]
hCV1247472	rs12551192	106.361322	0.043049	TSC1	0.23	Intronic	[A/G]
hCV2253563	rs633153	106.404371	0.011651	GFI1B	0.39	Intronic	[C/T]
hCV7582593	rs944204	106.416022	0.029705	GFI1B	0.45	Intergenic/Unknown	[C/G]
hCV2535358	rs623489	106.445727	0.051142	GTF3C5	0.33	Intergenic/Unknown	[C/T]

hCV2279860	rs685959	106.496869	0.018956	CEL	0.43	Intergenic/Unknown	[C/T]
hCV2535450	rs886017	106.515825	0.027814	RALGDS	0.48	Silent/UTR3	[A/G]
hCV2535940	rs671050	106.543639	0.029382	RALGDS	0.35	Intronic	[C/T]
hCV2535973	rs2073927	106.573021	0.009384	GBGT1/RALGDS	0.36	Intronic	[A/G]
hCV2980279	rs9411461	106.582405	0.037330	FS/GBGT1	0.36	Intergenic/Unknown	[C/G]
hCV2980256	rs10901243	106.619735	0.011154	OBP2B	0.30	Intergenic/Unknown	[C/G]
hCV27224742	N/A	106.630889	0.043658	OBP2B	0.23	Intergenic/Unknown	[C/T]
hCV3183098	rs2073824	106.674547	0.016859	ABO	0.38	Intronic	[A/G]
hCV3183164	rs529565	106.691406	0.059909	ABO	0.22	Intronic	[C/T]
hCV3183233	rs120858	106.751315	0.037673	SURF5/SURF6	0.48	Intronic	[A/G]
hCV3183190	N/A	106.788988	0.028952	SURF4	0.48	Intronic	[A/G]
hCV8784811	rs943623	106.817940	0.014386	XPMC2H/C9orf96	0.48	Intronic/UTR	[G/T]
hCV11572323	rs3118663	106.832326	0.029276	XPMC2H/ADAMTS13	0.48	Intronic	[A/G]
hCV3183371	rs652600	106.861602	0.013230	ADAMTS13	0.44	Intronic	[A/G]
hCV8784809	rs1055432	106.874832	0.080484	ADAMTS13/C9orf7	0.36	Silent Mutation	[A/G]
hCV2536686	N/A	106.955316	0.019154	ADAMTSL2	0.24	Intronic	[C/G]
hCV1247496	rs1105633	106.974470	0.050847	ADAMTSL2/KIAA0605	0.47	Intronic	[C/T]
hCV2971472	rs1029372	107.025317	0.016123	N/A	0.40	Intergenic/Unknown	[A/G]
hCV2535803	rs2519148	107.041440	0.014377	DBH	0.48	Intergenic/Unknown	[A/G]
hCV2535694	rs1611122	107.055817	0.013991	DBH	0.47	Intronic	[G/C]
hCV2535675	rs2073837	107.069808	0.025723	DBH/SARDH	0.24	Intronic	[A/G]
hCV11572672	rs1076149	107.095531	0.024936	SARDH	0.34	Intronic	[A/T]
hCV2540688	rs495464	107.120467	0.023127	SARDH	0.42	Intronic	[A/T]
hCV2537353	rs916620	107.143594	0.792522	SARDH	0.41	Intronic	[C/T]
hCV12020823	rs1980852	107.936116	0.974142	hCG1814720	0.50	Intergenic/Unknown	[A/G]
hCV3241385	rs968569	108.910258	0.993370	MGC29761/MRPS2	0.49	Intronic	[G/T]
hCV12019285	rs6563	109.903628	1.001538	NOTCH1	0.49	UTR3	[A/G]
hCV469299	N/A	110.905166	0.632732	FLJ31318	0.43	Intronic	[A/G]
hCV247127	rs2229948	111.537898		CACNA1B	0.33	UTR3/Silent	[A/G]

D9S934	91.7434 - 91.743612			Microsatellite	
D9S282	97.456102 - 97.456335			Microsatellite	
D9S2157	106.577163 - 106.577430			Microsatellite	

Table 4. Chromosome 13q SNPs

Celera ID	Public ID	Celera Mb Location	Mb Interval	Gene	MAF	SNP Type	SNP
hCV3077848	rs1107987	0.815377	1.047867	TUBA2	0.44	Intronic	[T/C]
hCV1813046	rs945367	1.863244	0.866118	GJB6	0.44	Intronic	[A/T]
hCV9183986	N/A	2.729362	0.664500	hCG1642913/1648309	0.40	Intergenic/Unknown	[C/T]
hCV2728143	rs1120978	3.393862	1.467574	hCG2019553	N/A	Intronic	[T/G]
hCV2702334	N/A	4.861436	1.049561	SGCG	0.50	Intronic	[A/C]
hCV265808	rs2096083	5.910997	1.170791	SPATA13	0.48	Intronic	[A/G]
hCV3126647	rs1924773	7.081788	0.818030	N/A	0.47	Intergenic/Unknown	[G/T]
hCV1906833	N/A	7.899818	1.166212	CDK8	0.44	Intronic	[A/G]
hCV8692413	rs1467591	9.066030	0.994051	GTF3A	0.49	Intergenic/Unknown	[A/G]
hCV1926749	N/A	10.060081	0.861621	FLT1	0.42	Intronic	[C/G]
hCV2731686	rs1023166	10.921702	0.978440	hCG29154	0.42	Intronic	[C/T]
hCV2539773	rs594411	11.900142		MGC2599	0.44	Intronic	[A/T]

D13S175	1.910207 - 1.910357					Microsatellite	
D13S629	11.667772 - 11.668212					Microsatellite	

Table 5. Chromosome 16q SNPs

Celera ID	Public ID	Celera Mb Location	Mb Interval	Gene	MAF	SNP Type	SNP
hCV2847280	N/A	53.348581	1.106331	CDH1	0.49	Intronic	[G/T]
hCV1430596	N/A	54.454912	1.206670	WWP2	0.44	Intronic	[G/T]
hCV11513151	N/A	55.661582	0.787548	N/A	0.43	Intergenic/Unknown	[G/T]
hCV11439127	rs1050363	56.449130	0.946085	FLJ20511/DHX38	0.48	Silent Mutation	[C/T]
hCV2191548	N/A	57.395215	1.144460	ATBF1	0.42	Intronic	[A/G]
hCV2850553	rs3325	58.539675	1.093499	N/A	0.45	Intergenic/Unknown	[A/G]
hCV2845735	rs3743607	59.633174	0.978347	CFDP1	0.42	Intronic	[C/T]
hCV7606101	rs1559330	60.611521	0.928785	CASPR4	0.48	Intergenic/Unknown	[C/T]
hCV2852259	rs2343039	61.540306	0.902992	HSRG1	0.44	Intronic	[G/T]
hCV11517904	rs2287972	62.443298	1.027547	N/A	0.44	Intronic	[C/T]
hCV8902185	rs1111230	63.470845	1.087958	N/A	0.49	Intergenic/Unknown	[A/G]
hCV1877093	N/A	64.558803	1.047647	hCG2040751	0.47	Intergenic/Unknown	[A/C]
hCV489829	N/A	65.606450	0.796298	BCMO1	0.45	Intronic	[A/T]
hCV1396200	rs4398102	66.402748	1.018372	HSD17B2	0.48	Intronic	[C/G]
hCV8092354	N/A	67.421120	1.001453	CDH13	0.46	Intronic	[A/G]
hCV1519129	rs2875857	68.422573	0.991637	MBTPS1	0.44	Intronic	[A/C]
hCV1430791	rs2291967	69.414210	0.841532	KIAA0513	0.50	Intronic	[G/T]
hCV3189974	rs2280378	70.255742	1.141016	ICSBP1	0.47	Intronic	[A/G]
hCV2925928	N/A	71.396758	0.995638	N/A	0.49	Intergenic/Unknown	[G/T]
hCV219649	N/A	72.392396		BANP	0.42	Intronic	[C/G]

D16S752	55.650714 - 55.650824					Microsatellite	
D16S516	63.423868 - 63.424034					Microsatellite	
D16S539	~ 70.689000					Microsatellite	

Table 6. Chromosome 18p SNPs

Celera ID	Public ID	Celera Mb Location	Mb Interval	Gene	MAF	SNP Type	SNP
hCV1367907	rs585578	0.079384	0.711168	USP14/THOC1	0.31	Intronic	[A/G]
hCV3013161	rs1608446	0.790552	0.968134	ADCYAP1	0.42	Intergenic/Unknown	[A/G]
hCV3235262	rs313021	1.758686	1.002189	N/A	0.46	Intergenic/Unknown	[A/G]
hCV3011859	rs598866	2.760875	1.105338	EMILIN-2	0.49	Intronic	[A/G]
hCV8121217	N/A	3.866213	1.251059	DLGAP1	0.40	Intronic	[G/T]
hCV11205167	N/A	5.117272	0.742332	MGC17515	0.49	Intergenic/Unknown	[T/C]
hCV1367564	rs736632	5.859604	1.144583	FLJ35936	0.45	Intronic	[C/T]
hCV3086198	rs2089760	7.004187	0.893609	LAMA1	0.47	Intergenic/Unknown	[C/T]
hCV1464431	N/A	7.897796	1.105522	PTPRM	0.50	Intronic	[A/G]
hCV7495855	rs977581	9.003318	0.845495	NDUFV2	0.41	Intronic	[G/T]
hCV612923	rs29062	9.848813	1.516311	VAPA	0.42	Intergenic/Unknown	[C/T]
hCV1637378	rs1026390	11.365124	0.794660	N/A	0.32	Intergenic/Unknown	[A/G]
hCV460143	rs4797665	12.159784	1.099948	CIDEA	0.40	Intergenic/Unknown	[A/G]
hCV8116885	rs872906	13.259732	0.721377	C18orf1	0.48	Intronic	[A/G]
hCV3047081	N/A	13.981109	1.478146	LOC162655	0.48	Intronic	N/A
hCV15827884	rs2127958	15.459255	0.581561	ROCK1	0.50	Intronic	[C/T]
hCV16100206	rs2847129	16.040816	0.945723	ABHD3	0.42	Intronic	[A/G]
hCV194412	N/A	16.986539	0.948744	hCG1643126	0.46	Intergenic/Unknown	[A/T]
hCV3204015	N/A	17.935283	0.970774	MIC1/NPC1	0.48	Intronic	[A/G]
hCV2019690	rs600958	18.906057	1.101806	hCG1643857	0.47	Intergenic/Unknown	[C/T]
hCV3253578	rs273756	20.007863	0.991459	hCG1656949	0.48	Intergenic/Unknown	[C/T]
hCV3004492	rs2438414	20.999322	0.838445	hCG38480	0.50	Intronic	[C/T]
hCV7492337	rs1467233	21.837767	1.206412	N/A	0.39	Intergenic/Unknown	[G/T]
hCV3117911	rs1354417	23.044179	0.734262	hCG1641502	0.49	UTR3	[A/G]
hCV11732665	N/A	23.778441		N/A	0.48	Intergenic/Unknown	[C/G]

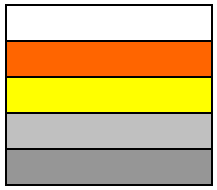
GATA166D05	1.998573-1.998875			Microsatellite	
D18S967	6.481905 - 6.482138			Microsatellite	
D18S843	8.498119 - 8.498307			Microsatellite	
D18S869	16.891612 - 16.891797			Microsatellite	

Table 7. Chromosome 19q SNPs

Celera ID	Public ID	Celera Mb Location	Mb Interval	Gene	MAF	SNP Type	SNP
hCV7610932	rs1035478	30.009929	0.944687	SLC7A9	0.49	Intronic	[C/T]
hCV2592102	rs285694	30.954616	1.282101	CHST8	0.50	Intronic	[C/G]
hCV25473593	N/A	32.236717	0.753525	KIAA1533/SCN1B/HPN	0.46	Intronic	[A/T]
hCV3111700	rs120960	32.990242	1.173085	SNX26	0.45	Intronic	[A/G]
hCV1277048	rs826285	34.163327	1.019939	N/A	0.41	Intergenic/Unknown	[A/T]
hCV1135244	rs953370	35.183266	0.815493	NYD-SP11	0.41	UTR3	[C/T]
hCV11507800	rs3745859	35.998759	1.028455	ACTN4/M9	0.48	Silent Mutation	[C/T]
hCV2275630	rs374185	37.027214	1.029339	CLC	0.42	Missense Mutation	[A/G]
hCV11465156	rs1870087	38.056553	1.009044	hCG20793	0.49	Intronic	[G/T]
hCV8597405	N/A	39.065597	1.401557	CEACAM6	0.42	Missense Mutation	[G/T]
hCV11975183	rs2024096	40.467154	0.753944	PSG5	0.38	Intergenic/Unknown	[C/G]
hCV2652664	rs417699	41.221098	0.987082	ZNF45	0.49	Nonsense Mutation	[A/G]
hCV3084818	rs760136	42.208180	0.935342	APOE	0.50	Intronic	[A/G]
hCV2884324	N/A	43.143522	0.875981	SPK	0.45	Intronic	[A/G]
hCV3168164	N/A	44.019503	1.039866	STRN4/PKD2	0.34	Intronic	[A/G]
hCV1996744	rs3786780	45.059369	1.013716	EHD2/GLTSCR2	0.50	Intronic	[G/C]
hCV2405293	rs602662	46.073085	1.041268	FUT2	0.47	Silent Mutation	[A/G]
hCV1844598	N/A	47.114353	0.888870	TSKS/CPT1C/PTOV1	0.49	Intronic	[C/T]
hCV1655600	rs1673028	48.003223	1.049302	MYBPC2	0.40	Intronic	[C/T]
hCV3057052	N/A	49.052525	1.149116	SIGLECL1	0.45	Missense Mutation	[A/G]
hCV3098516	rs619872	50.201641	1.128286	ZNF83	0.47	Intronic	[C/G]
hCV1997401	N/A	51.329927	0.950010	N/A	0.47	Intergenic/Unknown	[C/G]
hCV2996831	rs270790	52.279937	0.758504	KIR3DL7	0.45	Missense Mutation	[A/G]
hCV8879227	rs7478	53.038441	0.958148	KLP1/HSPC189	0.48	UTR3	[A/T]
hCV2658107	rs3760849	53.996589	0.917267	FLJ14011	0.42	Missense Mutation	[C/T]
hCV11701198	N/A	54.913856	1.076198	ZNF304/ZNF547	0.49	UTR3	[A/G]
hCV1116794	rs1465789	55.990054		ZNF132	0.48	Missense Mutation	[A/G]

D19S587	31.922494 - 31.922640					Microsatellite	
D19S211	40.184511 - 40.184704					Microsatellite	

D19S402	49.220919 - 49.221226			Microsatellite	



TaqMan® Assays-on-Demand℠
TaqMan® Assays-by-Design℠
Microsatellites
TaqMan® Assays-on-Demand℠ (First Group of Assays Added)
TaqMan® Assays-on-Demand℠ (Second Group of Assays Added)

APPENDIX C

Seven Candidate Regions for MS – Analysis Results

Table 1. Chromosome 1q Analysis

						MLOD			PDT		
SNP ID (Celera)	**SNP ID (dbSNP)**	**Mb (Celera)**	**Mb (dbSNP)**	**Gene**	**MAF**	**All**	**HLA All**	**HLA None**	**PDTSum**	**PDTAve**	**PDTGeno**
hCV1650709	rs734551	205.263268	228.301168	DISC1	0.46	0.00	0.03	0.02	0.648	0.700	0.740
hCV2798667	rs6665073	206.377503	229.413857	MGC13186	0.38	0.00	0.00	**1.15**	0.974	0.888	**0.038**
hCV8690690	rs701237	207.076236	230.112722	KCNK1	0.34	0.05	0.05	**1.13**	0.468	0.398	0.526
hCV2711850	rs564212	208.061974	231.098456	hCG1813648	0.38	0.74	0.87	0.52	0.739	0.562	0.841
hCV695532	rs423026	208.958848	232.041597	GNG4	0.49	0.02	0.01	0.58	0.786	0.705	0.132
hCV7540624	rs1266380	209.930048	233.007791	LGALS8	0.46	**1.55**	0.67	**1.73**	0.604	0.978	0.717
hCV16123987	rs2805432	210.934838	234.011263	RYR3	0.47	**2.60**	**2.00**	**1.51**	0.844	0.321	0.369
hCV401377	rs1557132	211.872092	234.944276	hCG23440	0.33	0.76	0.66	0.31	0.394	0.063	0.139
hCV2050524	rs2278644	213.123452	236.193833	hCG1984146	0.47	0.96	0.85	0.13	0.273	0.640	0.371
D1S1594		214.025			7 alleles	0.98	0.28	0.79	0.458	0.592	0.401
hCV12008520	rs1934338	214.026729	237.09664	FLJ21195	0.44	**1.47**	0.65	**1.80**	1.000	0.760	0.760
D1S547		215.010921			14 alleles	**2.07**	0.11	**1.89**	0.152	0.145	0.225
hCV605574	rs671989	215.031413	238.10244	CHML	0.48	**1.82**	0.65	0.76	1.000	0.857	0.879
D1S1634		~215.575			22 alleles	**2.67**	**1.24**	**1.06**	0.412	0.643	0.183
hCV26589316	N/A	215.794773	N/A	hCG1989348	0.44	**1.09**	0.41	0.13	0.860	0.769	0.393
hCV16244390	rs2490395	216.745721	239.784963	hCG16602	0.47	0.73	0.14	0.45	0.219	0.525	0.250
hCV9487525	rs1566661	218.184227	241.220314	hCG1660865	0.49	0.56	0.06	0.76	0.816	0.449	0.588
hCV11669158	rs9919234	218.991813	242.030031	FLJ10157	0.44	0.92	0.30	0.60	0.454	0.676	0.164
hCV15761414	rs3007406	220.059643	243.098399	FLJ23001	0.42	**1.12**	0.45	0.02	0.254	**0.001**	0.456
hCV11669332	rs6676750	221.059607	244.086851	hCG1646765	0.47	**1.29**	0.70	0.01	0.884	0.619	0.547
hCV12010310	rs2039824	221.797152	244.838678	hCG1724360	0.40	0.24	0.00	0.92	0.688	0.497	0.953

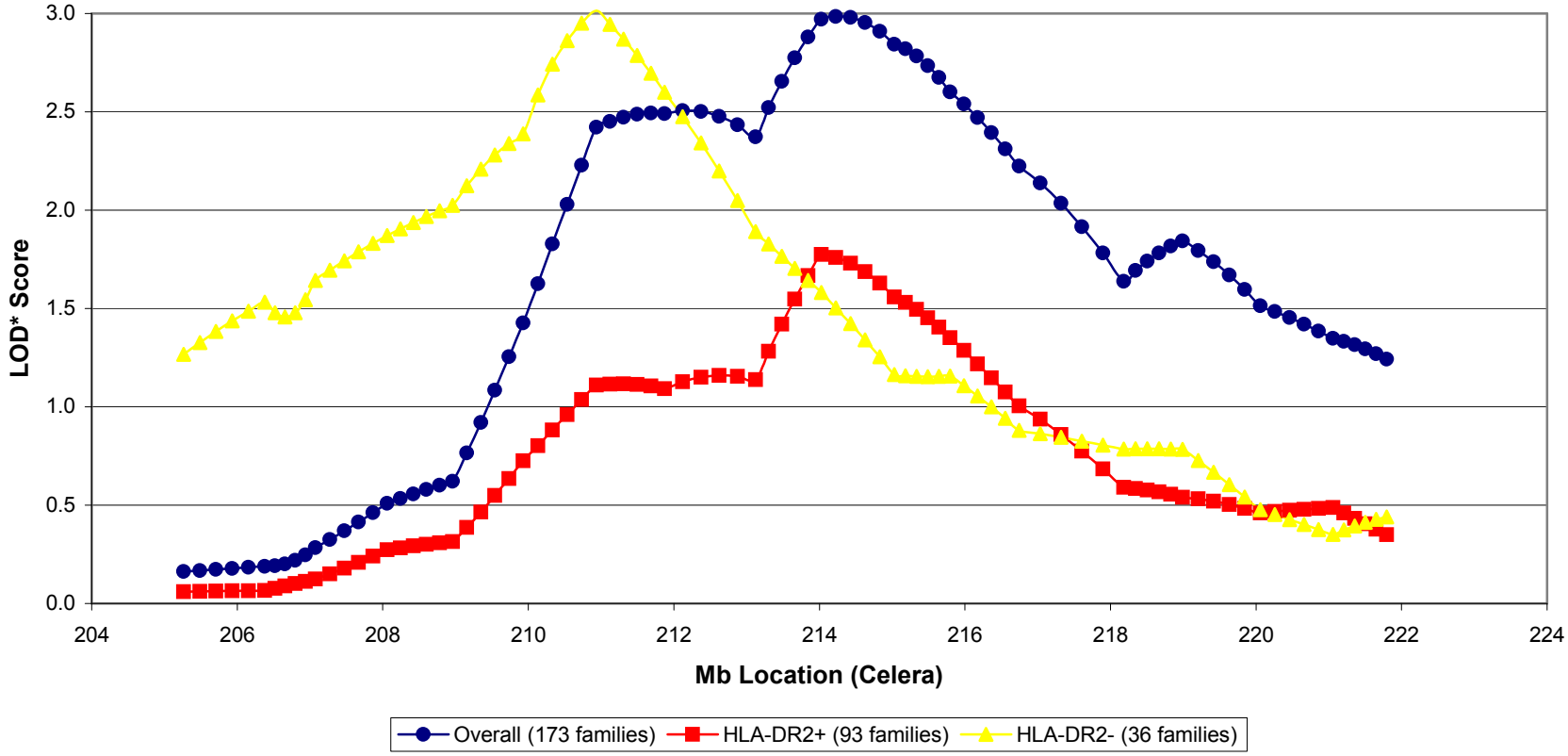


Figure 1. Chromosome 1q Multipoint Linkage Plot

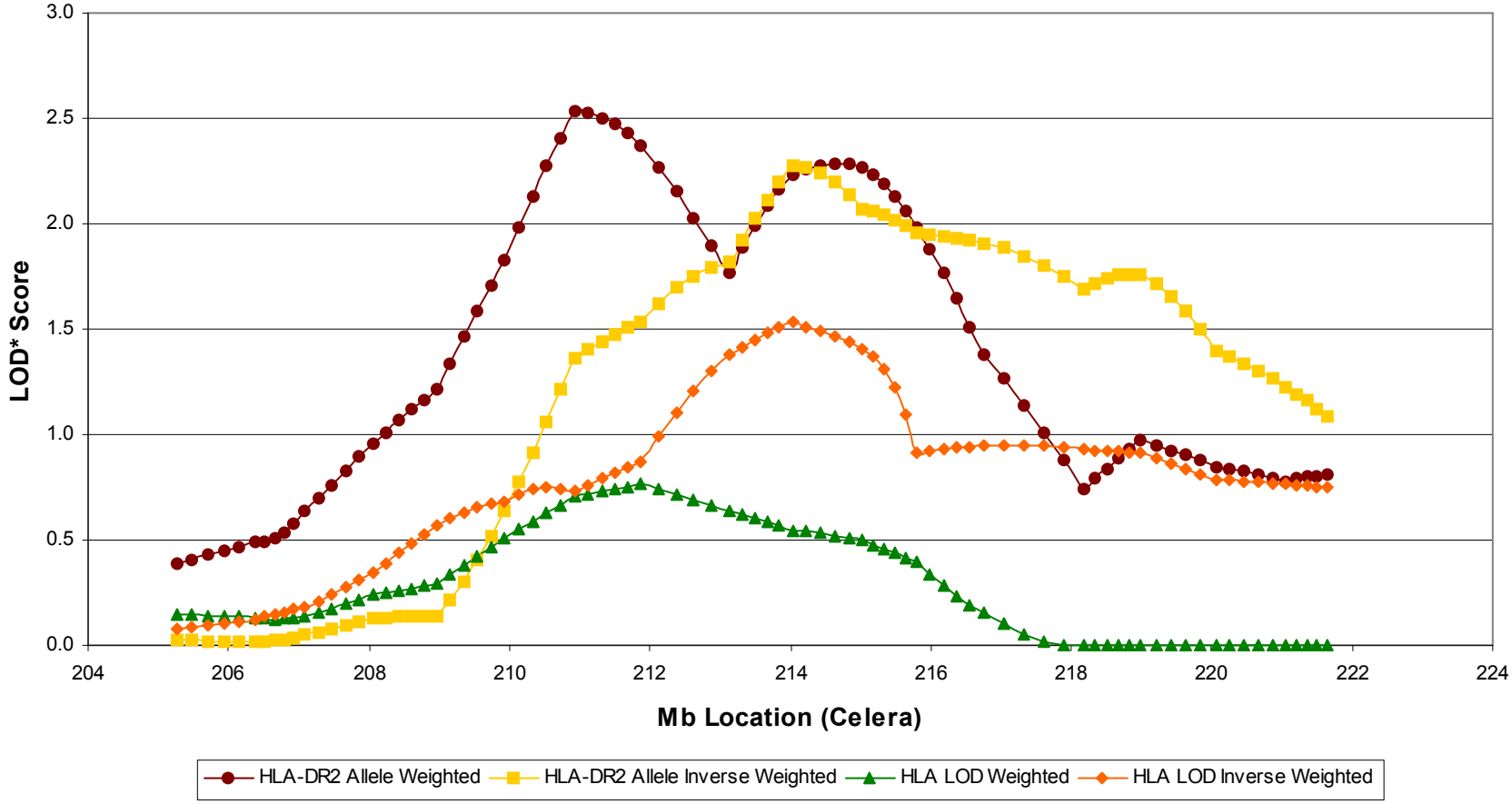


Figure 2. Chromosome 1q Multipoint Linkage Plot – HLA Conditional Analysis

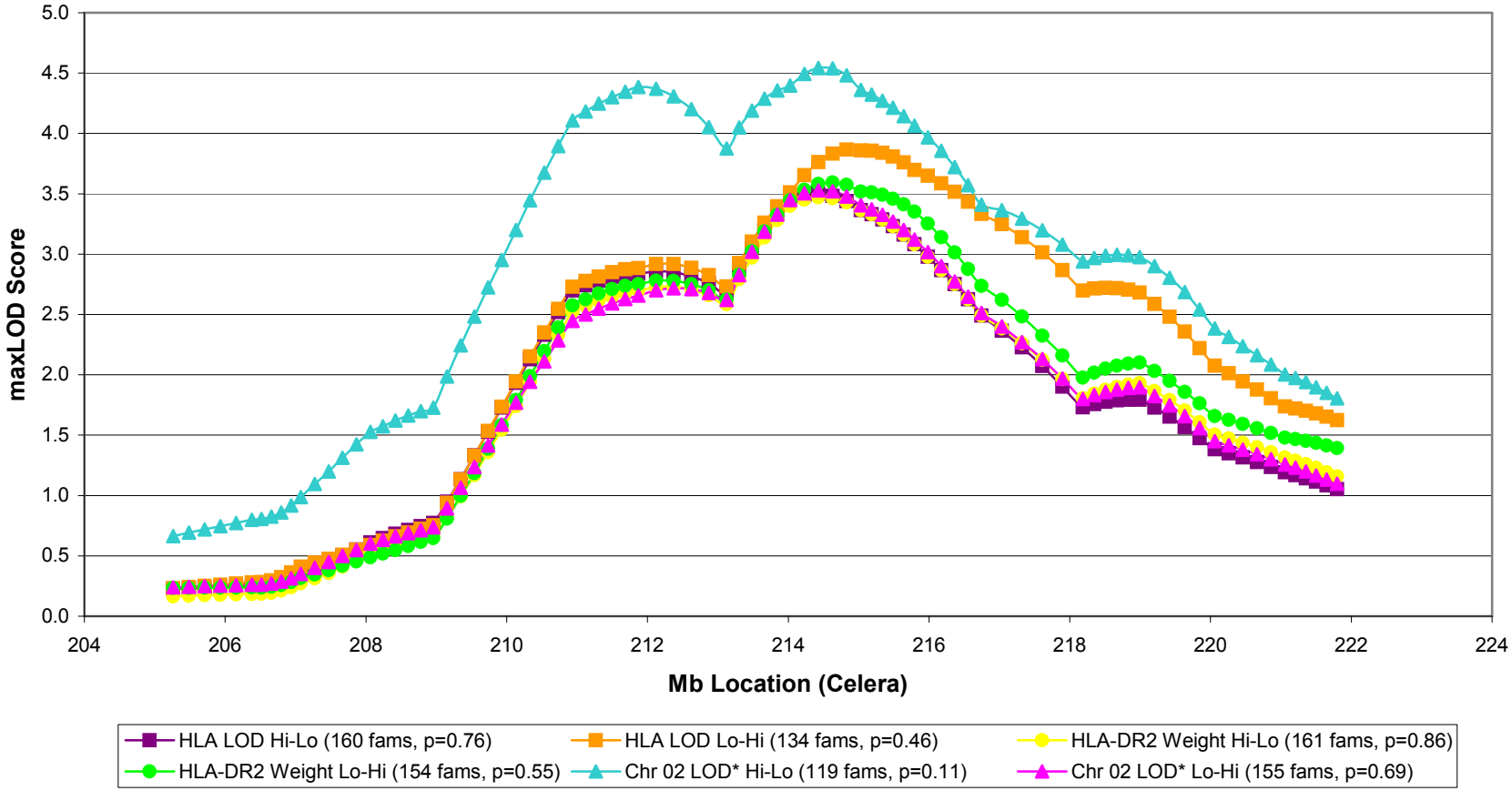


Figure 3. Chromosome 1q Multipoint Linkage Plot – Ordered Subset Analysis

The scale of the y-axis was changed from 3.0 to 5.0 to accommodate higher LOD scores.

Table 2. Chromosome 2q Analysis

						MLOD			PDT		
SNP ID (Celera)	**SNP ID (dbSNP)**	**Mb (Celera)**	**Mb (dbSNP)**	**Gene**	**MAF**	**All**	**HLA All**	**HLA None**	**PDTSum**	**PDTAve**	**PDTGeno**
hCV11827407	rs1982131	172.460711	178.683378	PDE11A	0.46	0.45	0.00	0.73	0.200	0.467	0.303
hCV2152432	rs2847	173.564451	179.795560	SESTD1	0.37	**1.21**	0.15	**1.62**	0.271	0.392	**0.046**
hCV31157131	rs13391691	174.744851	180.976047	N/A	0.38	0.06	0.00	0.00	0.241	0.796	0.519
hCV1051244	rs720453	175.747707	181.977337	N/A	0.46	0.00	0.00	0.00	0.321	0.625	0.622
hCV2074867	rs2128043	176.712645	182.942406	PDE1A	0.43	0.00	0.00	0.00	0.163	0.226	0.520
hCV1543267	rs2952363	177.598047	183.827451	LOC129401/NUP35	0.43	0.14	0.00	0.07	0.089	0.097	0.264
hCV426462	rs10497643	178.648492	184.878288	N/A	0.47	0.05	0.00	0.02	0.775	0.760	0.968
hCV7617345	rs889909	179.761231	185.990986	N/A	0.44	0.15	0.03	0.01	0.242	0.259	0.467
hCV11518651	rs10177153	180.818628	187.048301	N/A	0.34	0.20	0.14	0.00	0.427	0.810	0.204
hCV1590313	rs1528233	181.803615	188.032751	CALCRL	0.34	0.01	0.00	0.03	0.448	0.513	0.641
hCV474783	rs11692963	182.702316	188.932082	N/A	0.50	0.46	0.00	**1.03**	0.792	0.568	0.320
hCV1729153	rs3134646	183.474639	189.705349	COL3A1	0.48	0.00	0.00	0.42	0.268	0.862	0.375
hCV2831378	rs785260	184.615046	190.845519	N/A	0.48	0.00	0.00	0.16	0.335	0.540	0.486
hCV22274194	rs3771300	185.430195	191.661102	STAT1;GLS	0.50	0.04	0.00	0.17	0.608	0.374	0.860
hCV16207964	rs2356955	186.630420	192.860529	TMEFF2	0.48	0.51	0.00	**1.42**	0.728	0.485	0.358
hCV541573	rs717621	187.036077	193.266456	N/A	0.49	0.03	0.00	0.80	0.746	0.866	0.880
hCV11950973	rs726129	187.898443	194.129216	N/A	0.33	0.00	0.00	0.16	0.752	0.508	**0.007**
hCV1346631	rs4591357	188.834786	195.073110	hCG2038662	0.28	0.00	0.00	0.05	0.310	0.942	0.576
hCV8744355	rs1019845	189.705845	195.774199	Unknown	0.45	0.02	0.00	0.21	0.828	0.942	0.310
hCV7620124	rs1551838	190.483155	196.552819	DNAH7	0.40	0.14	0.05	0.21	0.825	0.346	0.935
hCV1291059	rs10931768	191.612951	197.682576	LOC91526	0.45	0.03	0.00	0.29	0.244	0.371	0.140
hCV2123977	rs1064213	192.706007	198.775746	PLCL1	0.46	0.16	0.00	0.52	0.543	0.869	0.618
GATA149B10		193.459832			7 alleles	0.00	0.00	0.00	0.527	0.129	0.431
hCV2153217	rs2881208	193.893141	199.963114	hCG20092	0.37	0.87	0.05	**1.18**	0.275	0.882	0.124
hCV3230597	rs1124639	194.530873	200.601250	FLJ38973	0.49	0.00	0.00	0.41	0.264	0.478	0.459
hCV1223380	rs11683632	195.395454	201.466080	hCG1811467	0.45	0.17	0.01	0.10	0.154	**0.040**	0.395
hCV347110	rs759419	196.518423	202.592153	ALS2CR7	0.47	0.53	0.00	**1.48**	0.248	**0.053**	0.607
hCV8761981	rs3845802	197.493359	203.566178	WDR12/ALS2CR14	0.44	0.00	0.00	0.00	0.300	0.736	0.127
hCV2821059	rs926169	198.477573	204.548258	CTLA4	0.39	0.08	0.00	0.10	0.100	0.273	0.289

D2S1384		198.981939			10 alleles	0.00	0.00	**1.12**	0.354	0.799	0.436
hCV1572836	rs7557452	199.681323	205.751906	ALS2CR19	0.44	0.00	0.00	0.08	0.731	0.838	0.853
hCV2186447	rs15994	200.500000	206.487802	NRP2	0.36	0.08	0.00	0.28	0.814	0.637	0.734
hCV2772287	rs2287508	201.672078	207.736124	CPO	0.49	0.31	0.00	0.55	0.642	0.451	0.907
hCV3090677	rs4675713	202.403786	208.467414	FZD5	0.45	0.18	0.00	0.18	0.339	0.408	0.596
hCV1737148	rs12151408	203.284385	209.342065	hCG2041988	0.42	0.07	0.00	**1.02**	0.246	0.158	0.298
hCV2668266	rs731953	204.443380	210.501076	hCG2040260	0.41	0.13	0.04	0.69	0.978	0.646	0.992
hCV2034024	rs2887914	205.260989	211.318613	CPS1/PRO0132	0.44	0.05	0.00	0.76	0.225	0.687	0.534
D2S1365		205.839346			10 alleles	0.00	0.00	0.65	0.652	0.606	0.718
hCV8835839	rs714393	206.463661	212.524224	ERBB4	0.47	0.07	0.00	0.12	0.904	0.858	0.250
hCV1861868	rs1871946	207.643260	213.704913	ZNFN1A2	0.42	0.00	0.00	0.01	0.238	0.162	0.367
hCV1552068	rs2372109	208.620868	214.678290	N/A	0.41	0.02	0.00	**1.68**	0.518	0.277	0.209

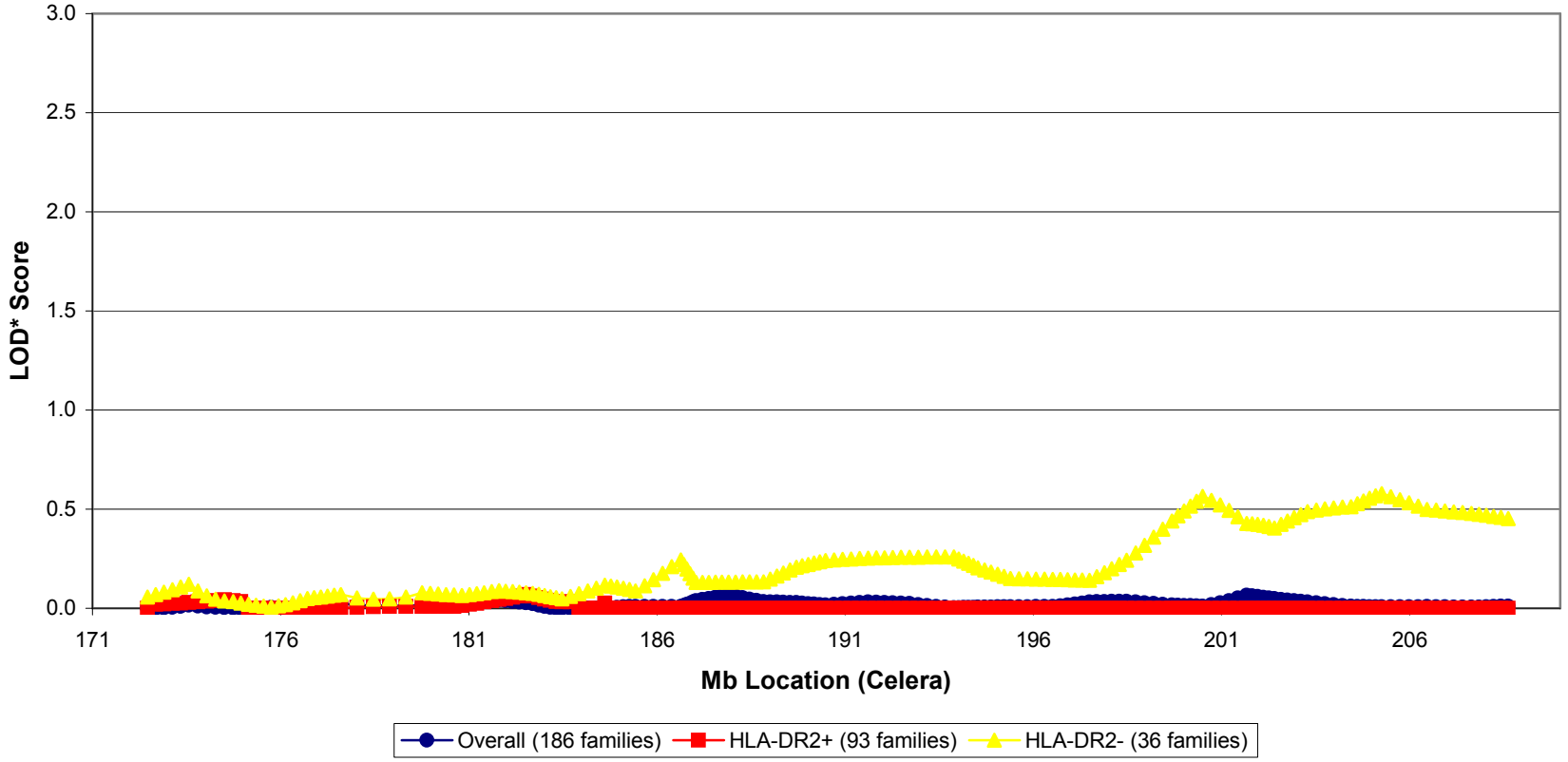


Figure 4. Chromosome 2q Multipoint Linkage Plot

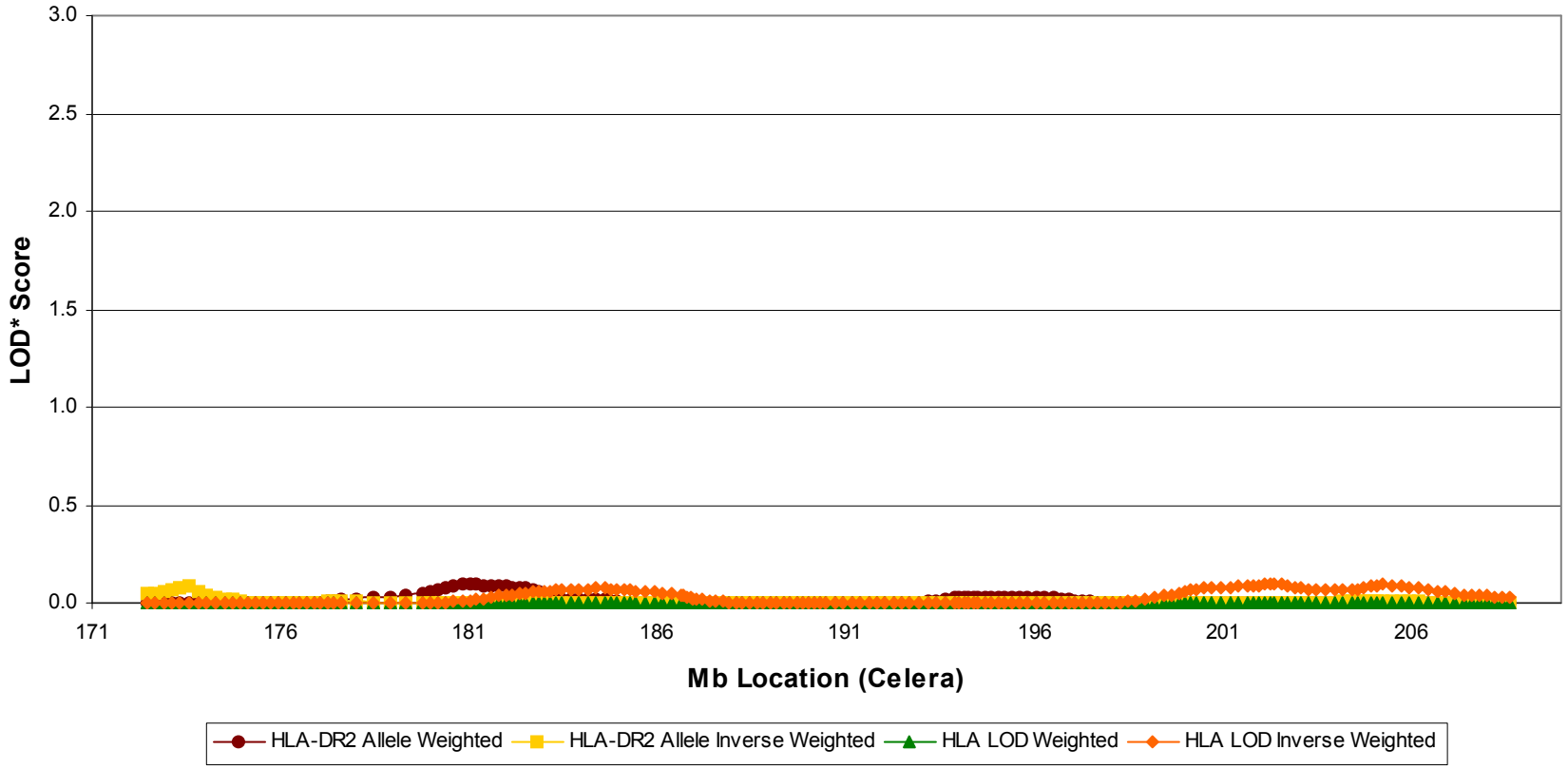


Figure 5. Chromosome 2q Multipoint Linkage Plot – HLA Conditional Analysis

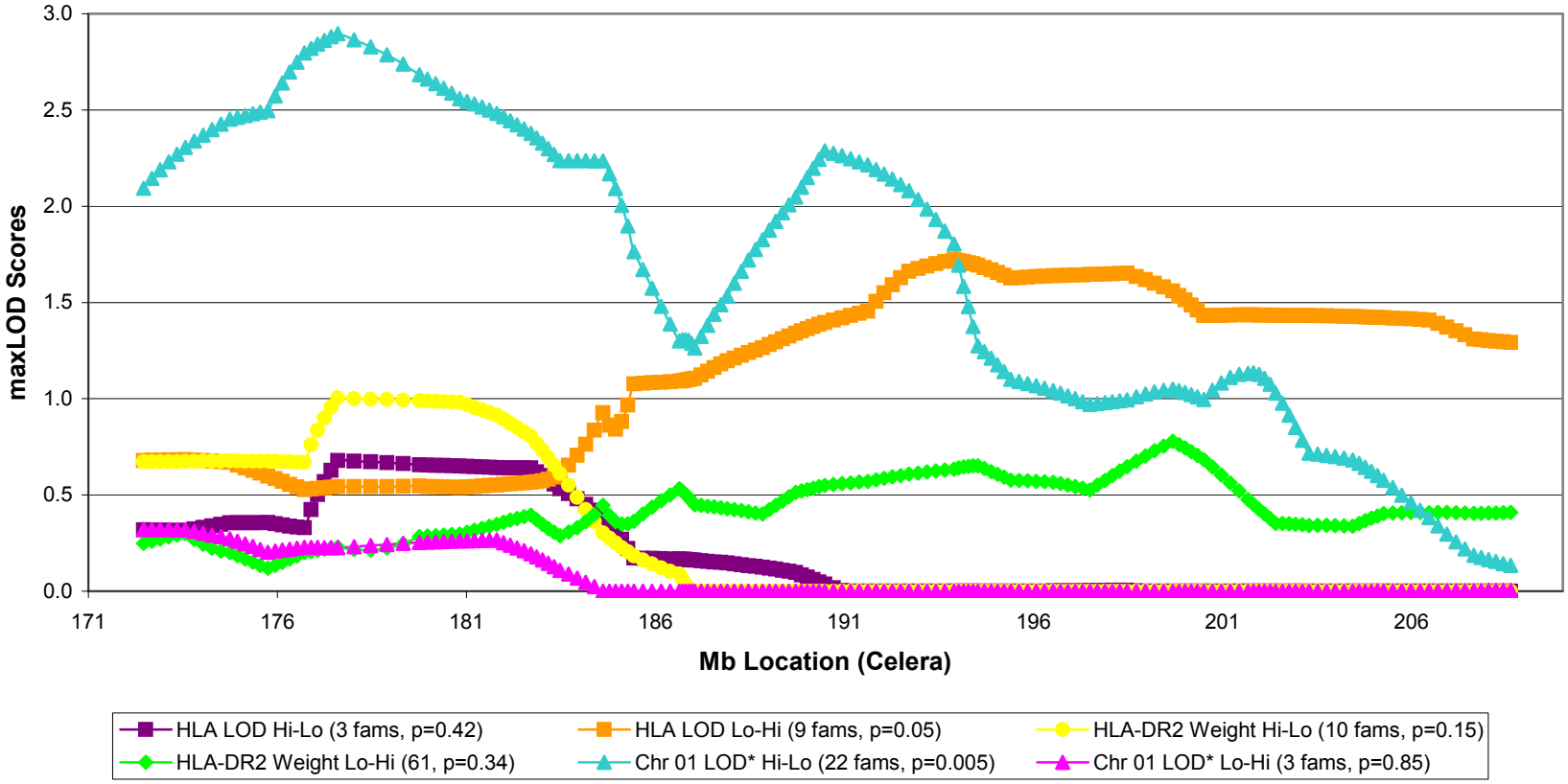


Figure 6. Chromosome 2q Multipoint Linkage Plot – Ordered Subset Analysis

Table 3. Chromosome 9q Analysis

						MLOD			PDT		
SNP ID (Celera)	SNP ID (dbSNP)	Mb (Celera)	Mb (dbSNP)	Gene	MAF	All	HLA All	HLA None	PDTSum	PDTAve	PDTGeno
hCV1899341	rs3810928	87.463072	113.895331	AMBP	0.48	0.07	0.00	0.00	0.169	0.667	0.401
hCV3022586	rs1061494	88.494049	114.926134	TNC	0.40	0.00	0.02	0.00	0.254	0.291	0.531
hCV7593836	rs971037	89.314153	115.746096	EST-YD1	0.42	0.00	0.00	0.00	0.812	0.231	0.772
hCV1979634	rs10733620	90.172104	116.604046	ASTN2	0.38	0.00	0.02	0.00	0.476	0.790	0.416
hCV11722141	rs1927911	91.117821	117.549608	TLR4	0.23	0.70	0.96	0.09	0.152	0.108	0.451
D9S934		91.743400			9 alleles	0.23	0.17	0.07	0.135	0.443	0.336
hCV1920588	rs1324623	92.767997	119.199751	DBCCR1	0.35	0.15	0.03	0.77	0.679	0.927	0.940
hCV1219009	rs3747850	94.713428	121.144482	GSN	0.49	0.00	0.00	0.02	0.686	0.157	0.696
hCV11884087	rs4679	95.562902	121.994167	NDUFA8	0.42	0.25	0.14	0.21	0.317	0.421	0.130
hCV58657	rs12686320	96.805159	123.236485	KIAA1608	0.37	0.12	0.06	0.20	0.124	0.420	0.393
hCV8780788	rs1042486	97.430687	123.862987	LHX2	0.48	0.00	0.00	0.02	0.079	**0.003**	0.293
D9S282		97.456102			9 alleles	0.25	0.00	0.35	0.377	0.631	0.745
hCV302240	rs501963	98.610850	125.04247	RAB9P40	0.42	**1.89**	0.28	**1.47**	0.214	0.315	0.216
hCV8782473	N/A	99.379990	125.805096	PBX3	0.29	0.00	0.11	0.01	**0.039**	0.346	0.132
hCV2700890	rs874799	100.376610	126.801559	RALGPS1A	0.49	0.38	0.07	0.00	0.593	0.470	0.219
hCV580692	rs514024	101.154940	127.583624	SH2D3C	4 alleles	0.00	0.00	0.00	0.905	0.921	0.397
hCV16180096	rs2273866	102.354750	128.782445	KIAA1094	0.39	0.00	0.00	0.06	0.771	0.684	0.866
hCV3180154	rs2296793	103.236092	129.664612	DYT1	0.25	0.95	0.00	**1.81**	0.571	0.516	0.140
hCV2605168	rs1056171	104.299288	130.790555	ABL1	0.48	0.02	0.00	0.12	0.096	0.757	0.187
hCV8782344	rs1056899	105.681837	132.169455	KIAA0625	0.29	0.26	0.14	0.05	0.125	0.488	0.365
D9S2157		106.577163			11 alleles	**2.65**	**1.84**	0.28	0.115	0.658	0.157
hCV3183190	rs1179037	106.788988	133.268063	SURF4	0.46	**1.48**	0.98	0.02	0.623	0.224	0.879
hCV12020823	rs1980852	107.936116	134.647976	hCG1814720	0.39	**1.07**	0.07	**1.43**	0.088	0.181	0.300
hCV3241385	rs968569	108.910258	135.619073	MGC29761/MRPS2	0.49	0.83	0.75	0.26	0.282	0.608	0.449
hCV12019285	rs6563	109.903628	136.665021	NOTCH1	0.47	0.40	0.53	0.22	0.410	0.150	0.680

hCV469299	rs11137268	110.905166	137.656213	FLJ31318	0.41	0.07	0.02	0.07	0.334	0.682	0.634
hCV247127	rs2229948	111.537898	138.284713	CACNA1B	0.29	0.00	0.00	0.01	0.359	0.187	0.169

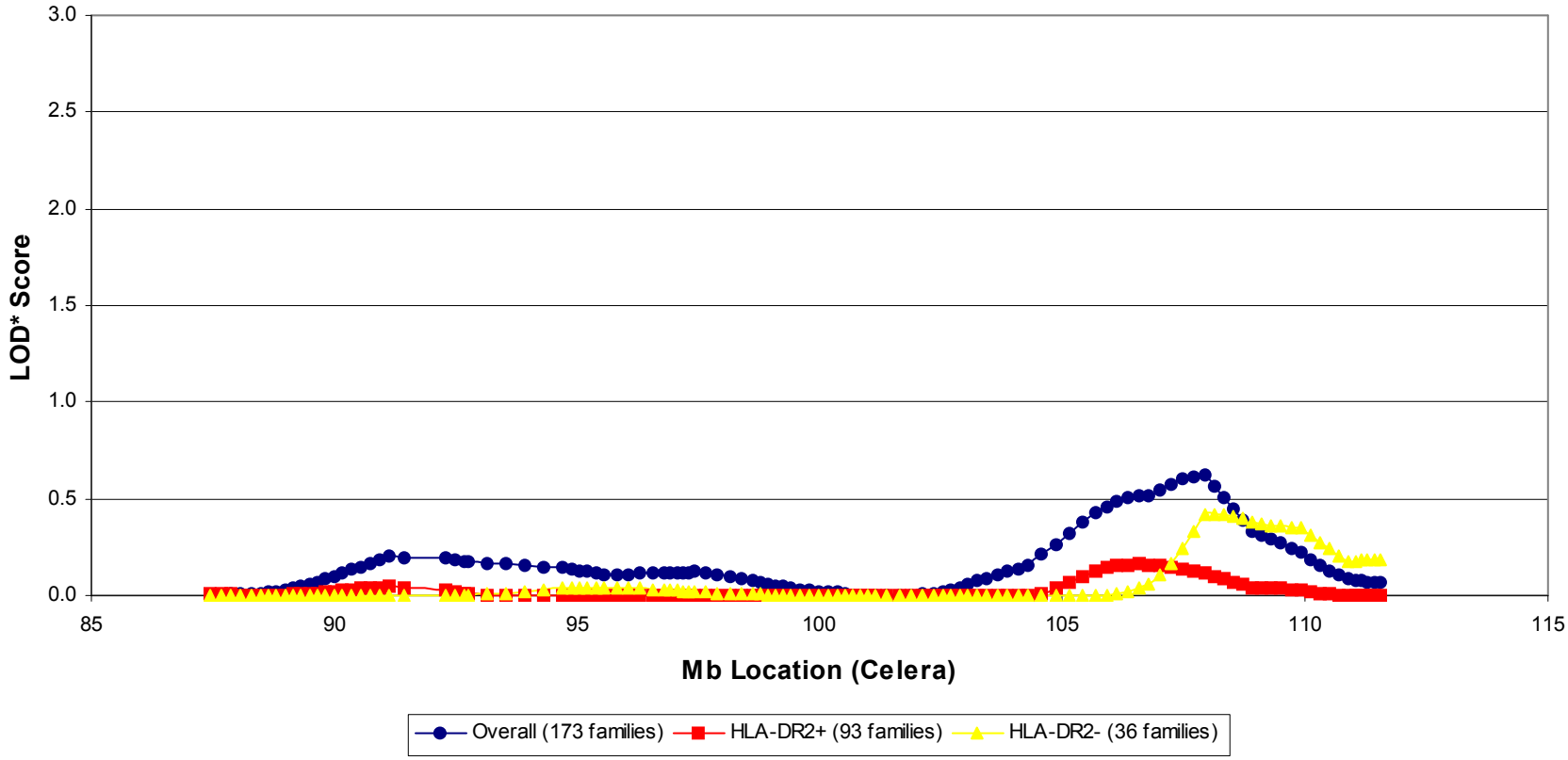


Figure 7. Chromosome 9q Multipoint Linkage Plot

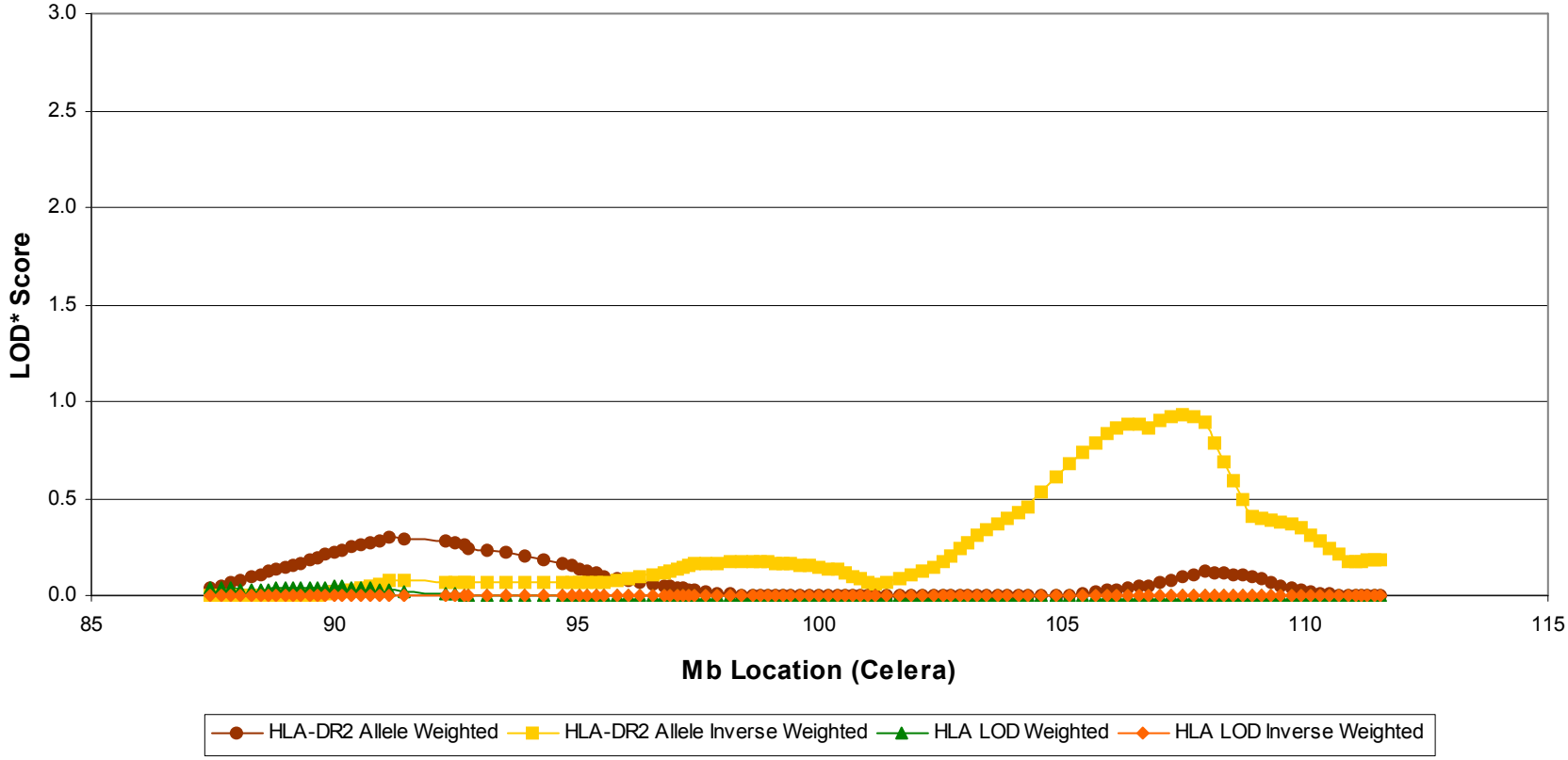


Figure 8. Chromosome 9q Multipoint Linkage Plot – HLA Conditional Analysis

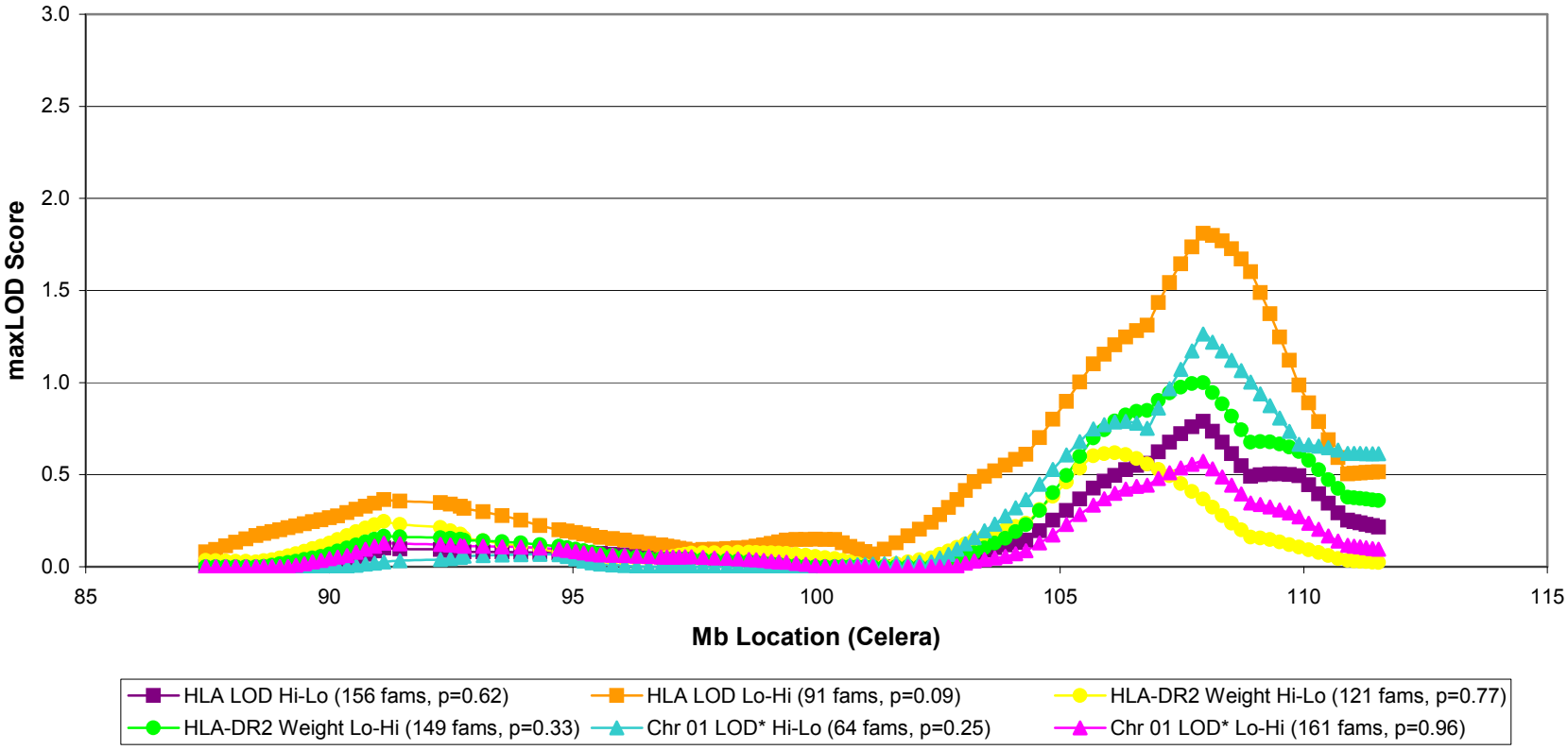


Figure 9. Chromosome 9q Multipoint Linkage Plot – Ordered Subset Analysis

Table 4. Chromosome 13q Analysis

						MLOD			PDT		
SNP ID (Celera)	SNP ID (dbSNP)	Mb (Celera)	Mb (dbSNP)	Gene	MAF	All	HLA All	HLA None	PDTSum	PDTAve	PDTGeno
hCV3077848	rs491873	0.815377	18.651438	TUBA2	0.41	0.68	0.60	0.20	0.634	0.963	0.579
hCV1813046	rs945367	1.863244	19.699425	GJB6	0.46	0.92	**1.36**	0.01	0.529	0.281	0.486
D13S175		1.910207			10 alleles	0.30	0.19	0.03	0.500	0.553	0.386
hCV9183986	rs9509528	3.393862	20.565388	hCG1642913	0.38	0.00	0.02	0.00	0.608	0.820	0.476
hCV2728143	rs1120978	3.393862	21.769874	hCG2019553	0.36	0.16	0.01	0.00	0.461	0.865	0.800
hCV2702334	rs1536723	4.861436	22.696431	SGCG	0.49	0.10	0.44	0.00	0.673	0.637	0.855
hCV265808	rs2096083	5.910997	23.746336	SPATA13	0.45	0.00	0.00	0.00	0.132	0.412	0.241
hCV3126647	rs1924773	7.081788	24.915624	N/A	0.49	0.00	0.00	0.00	0.588	0.588	0.609
hCV1906833	rs9512166	7.899818	25.729717	CDK8	0.43	0.00	0.00	0.00	0.603	0.164	0.567
hCV8692413	rs1467591	9.066030	26.895352	GTF3A	0.46	0.03	0.02	0.28	0.365	0.624	0.890
hCV1926749	rs9551468	10.060081	27.883316	FLT1	0.47	0.00	0.00	0.00	0.146	**0.044**	0.091
hCV2731686	rs1023166	10.921702	28.743810	hCG29154	0.46	0.00	0.00	0.00	0.773	0.452	0.269
D13S629		11.667772			13 alleles	0.00	0.00	0.10	0.806	0.715	0.509
hCV2539773	rs594411	11.900142	29.728600	MGC2599	0.32	0.00	0.00	0.00	0.642	0.692	0.876

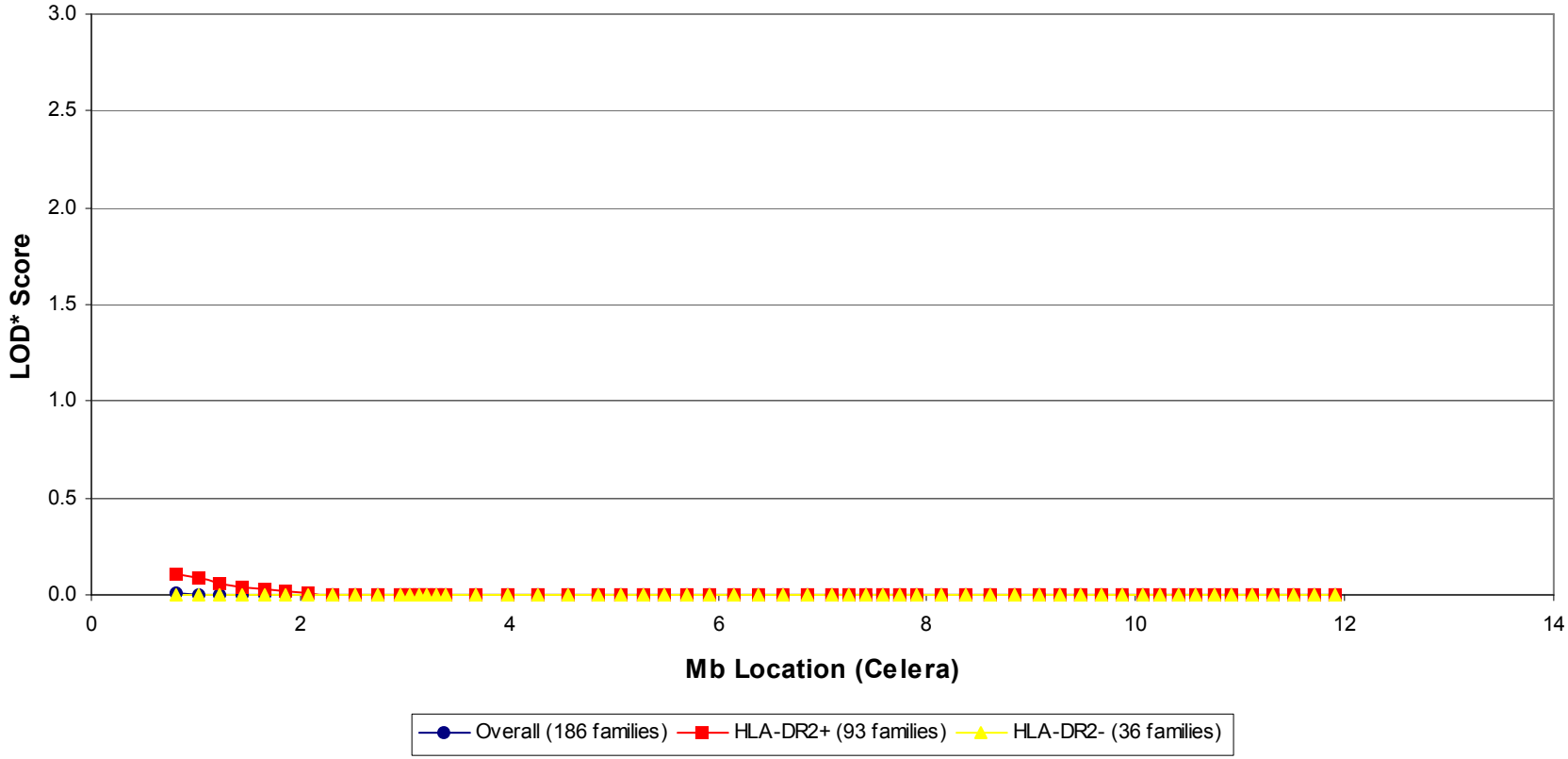


Figure 10. Chromosome 13q Multipoint Linkage Plot

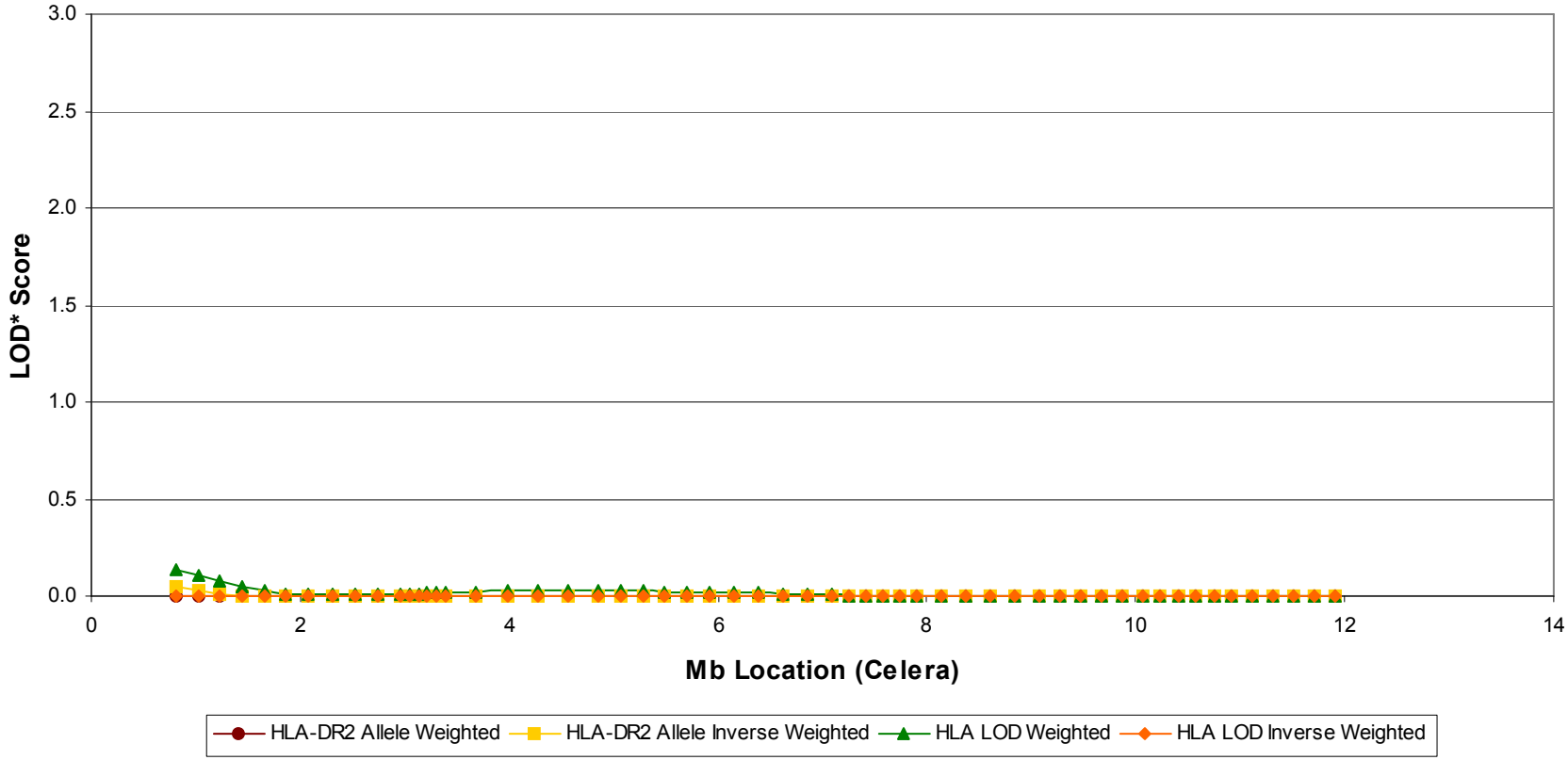


Figure 11. Chromosome 13q Multipoint Linkage Plot – HLA Conditional Analysis

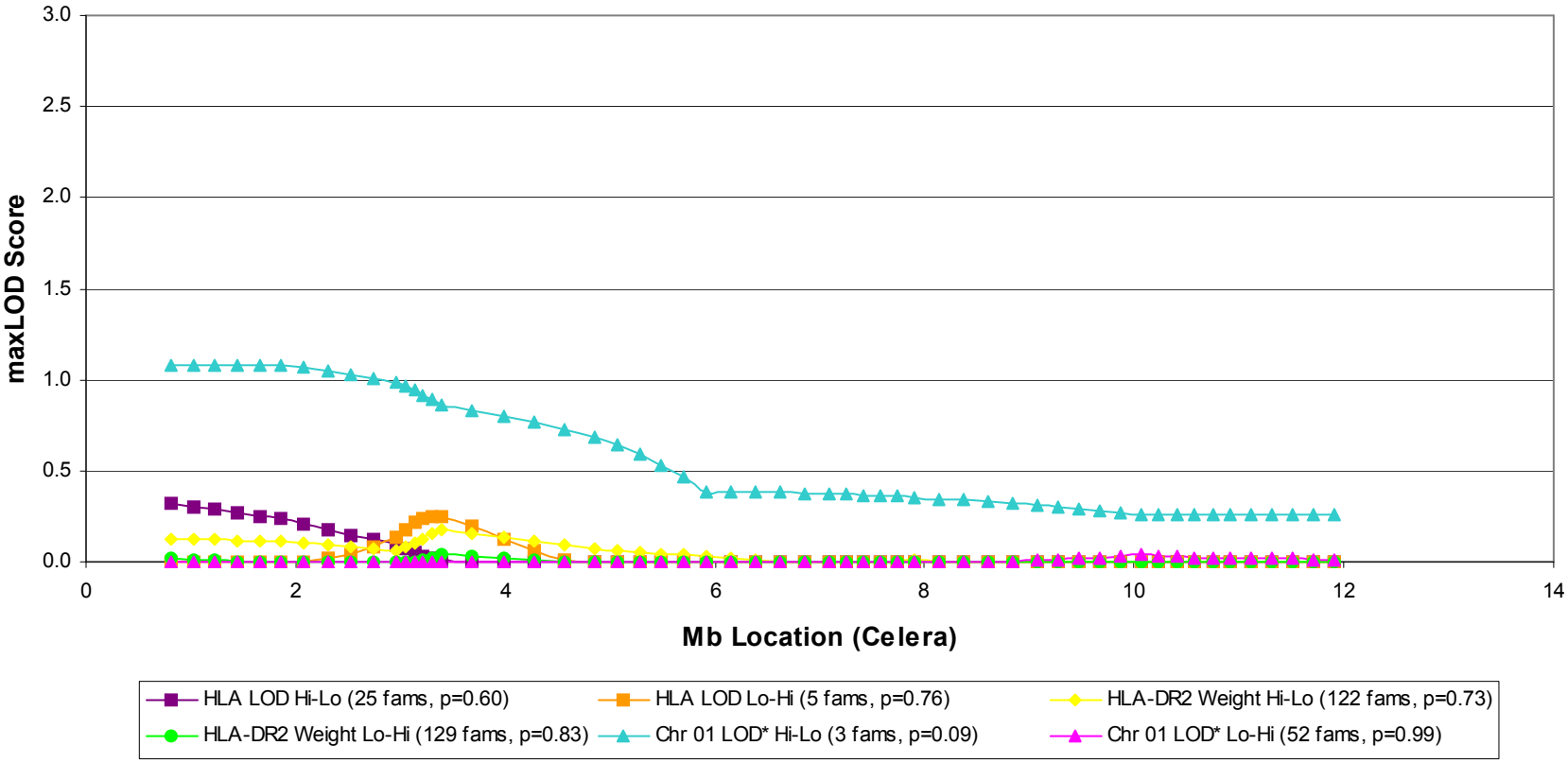


Figure 12. Chromosome 13q Multipoint Linkage Plot – Ordered Subset Analysis

Table 5. Chromosome 16q Analysis

						MLOD			PDT		
SNP ID (Celera)	**SNP ID (dbSNP)**	**Mb (Celera)**	**Mb (dbSNP)**	**Gene**	**MAF**	**All**	**HLA All**	**HLA None**	**PDTSum**	**PDTAve**	**PDTGeno**
hCV2847280	rs10431923	53.348581	67.396764	CDH1	0.50	0.02	0.08	0.00	0.587	0.663	0.702
hCV1430596	rs8049373	54.454912	68.497206	WWP2	0.38	0.04	0.33	0.22	0.048	0.109	0.180
D16S752		55.650714			7 alleles	0.00	0.34	0.00	0.863	0.978	0.636
hCV11513151	N/A	55.661582	69.903551	N/A	0.41	0.00	0.10	0.02	0.537	0.691	0.504
hCV11439127	rs1050363	56.449130	70.692515	FLJ20511/DHX38	0.47	0.00	0.08	0.00	0.664	0.347	0.209
hCV2191548	rs11641701	57.395215	71.636713	ATBF1	0.36	0.25	0.33	0.00	0.725	0.749	0.929
hCV2850553	rs3325	58.539675	72.783903	N/A	0.46	0.55	0.03	0.01	0.167	0.381	0.396
hCV2845735	rs3743607	59.633174	73.896356	CFDP1	0.38	0.20	**1.39**	0.01	0.143	0.869	0.187
hCV7606101	rs1559330	60.611521	74.866582	CASPR4	0.39	0.00	0.00	0.48	0.062	0.217	0.170
hCV2852259	rs2343039	61.540306	75.80028	HSRG1	0.48	0.20	0.00	0.03	0.694	0.082	0.842
hCV11517904	rs2287972	62.443298	76.701108	N/A	0.49	0.03	0.00	0.03	0.682	0.570	0.530
D16S516		63.423868			9 alleles	**1.32**	**1.37**	0.30	0.749	0.484	0.273
hCV8902185	rs1111230	63.470845	77.728638	N/A	0.49	0.34	0.28	0.28	0.318	0.877	0.107
hCV1877093	N/A	64.558803	78.817454	hCG2040751	0.48	0.99	**1.24**	0.11	0.495	0.533	0.831
hCV489829	rs7192170	65.606450	79.866614	BCMO1	0.45	**1.25**	**1.47**	0.40	0.439	0.202	0.192
hCV1396200	rs4398102	66.402748	80.661587	HSD17B2	0.45	0.85	0.56	0.32	0.482	0.671	0.711
hCV8092354	N/A	67.421120	816.81126	CDH13	0.49	0.00	0.00	0.87	0.793	0.732	0.641
hCV1519129	rs2875857	68.422573	82.678157	MBTPS1	0.45	0.00	0.00	0.08	0.949	0.580	0.572
hCV1430791	rs2291967	69.414210	83.669503	KIAA0513	0.47	0.16	0.31	0.00	0.324	0.249	0.668
hCV3189974	rs2280378	70.255742	84.510246	ICSBP1	0.44	0.00	0.00	0.56	0.691	0.104	0.498
D16S539		~70.689000			7 alleles	0.00	0.00	0.03	0.953	0.897	0.891
hCV2925928	rs11646219	71.396758	85.654078	N/A	0.46	0.00	0.03	0.02	0.947	0.594	0.948
hCV219649	rs12931579	72.392396	86.659431	BANP	0.36	0.00	0.02	0.02	0.975	0.420	0.960

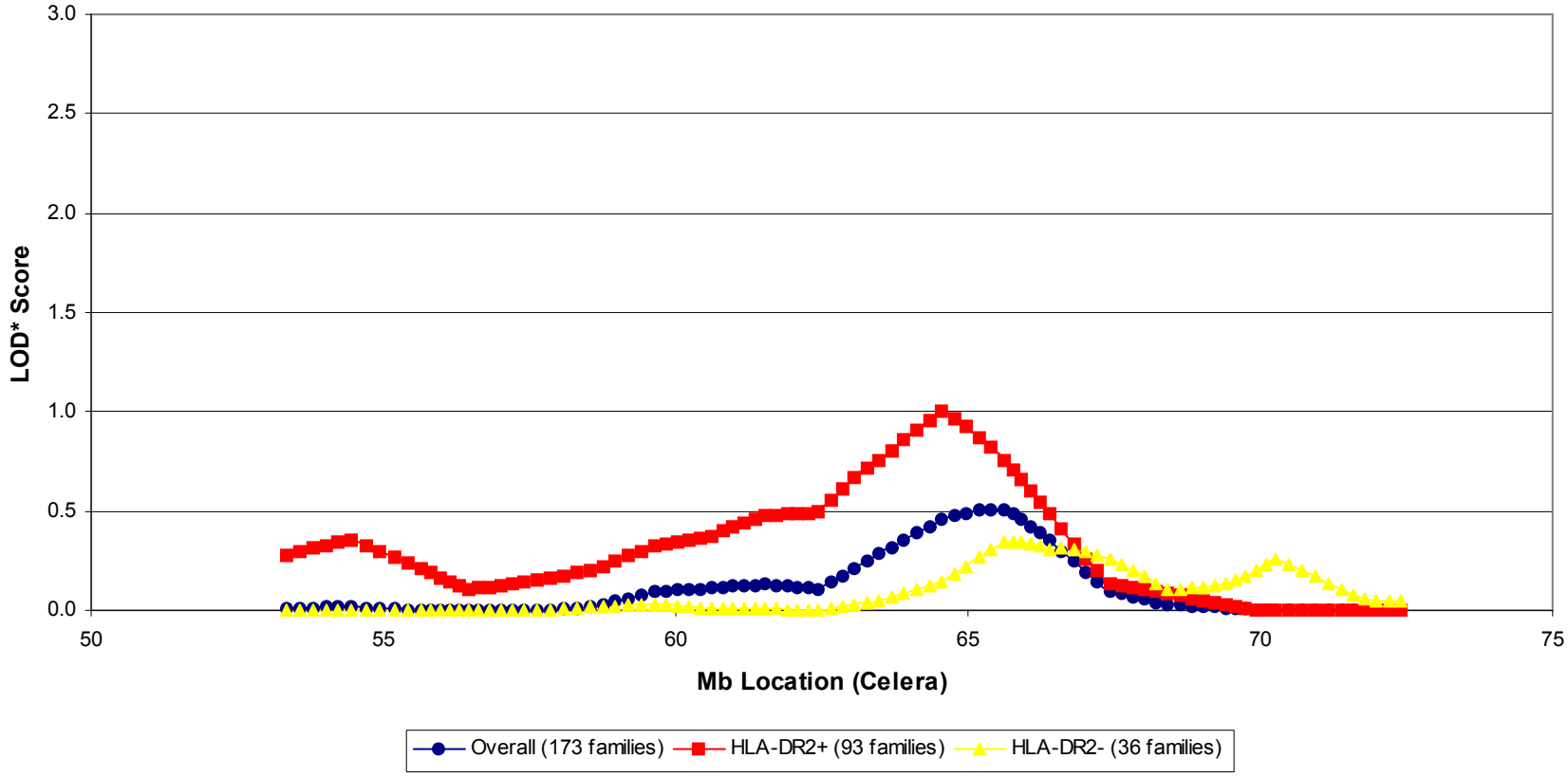


Figure 13. Chromosome 16q Multipoint Linkage Plot

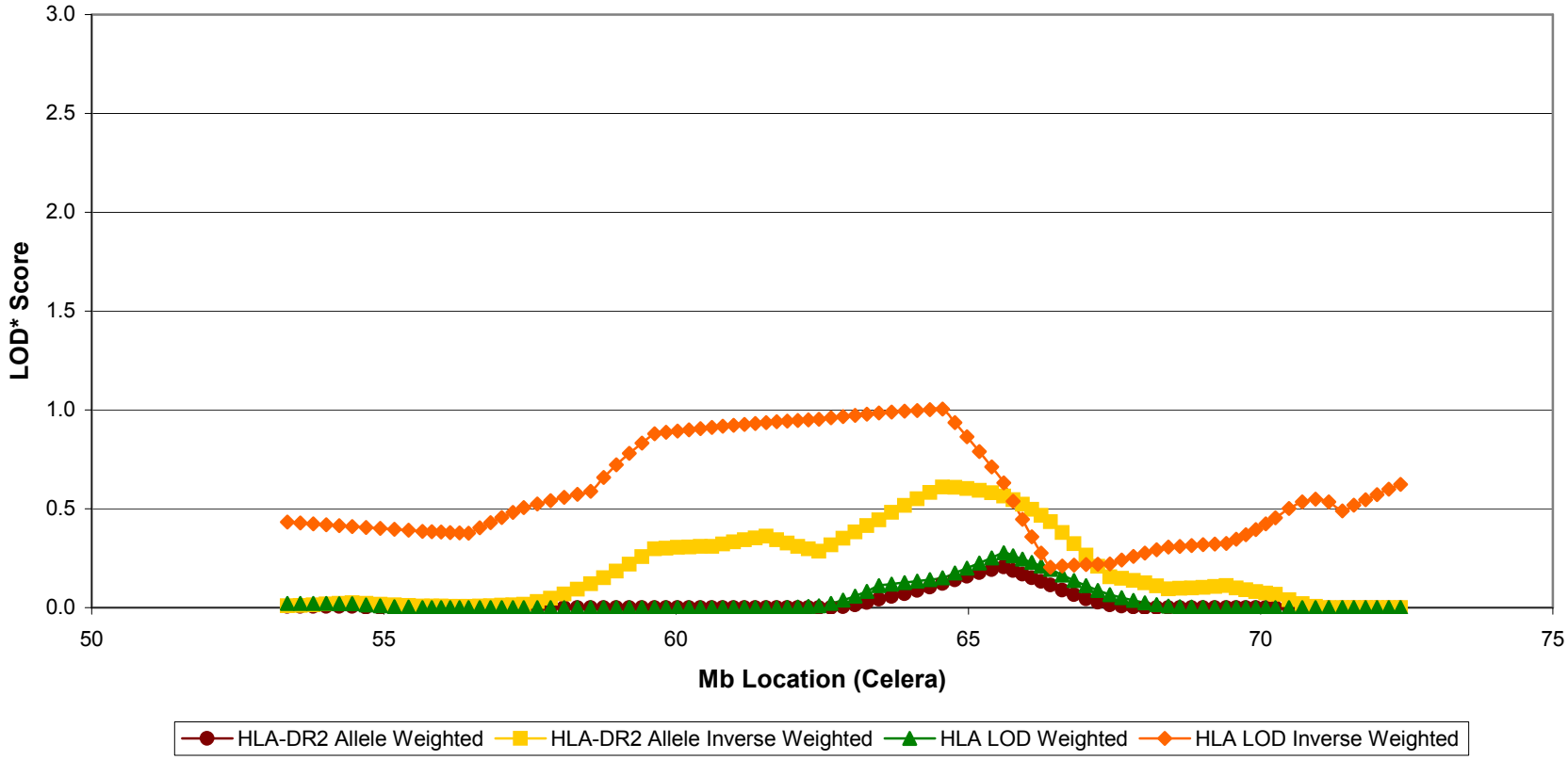


Figure 14. Chromosome 16q Multipoint Linkage Plot – HLA Conditional Analysis

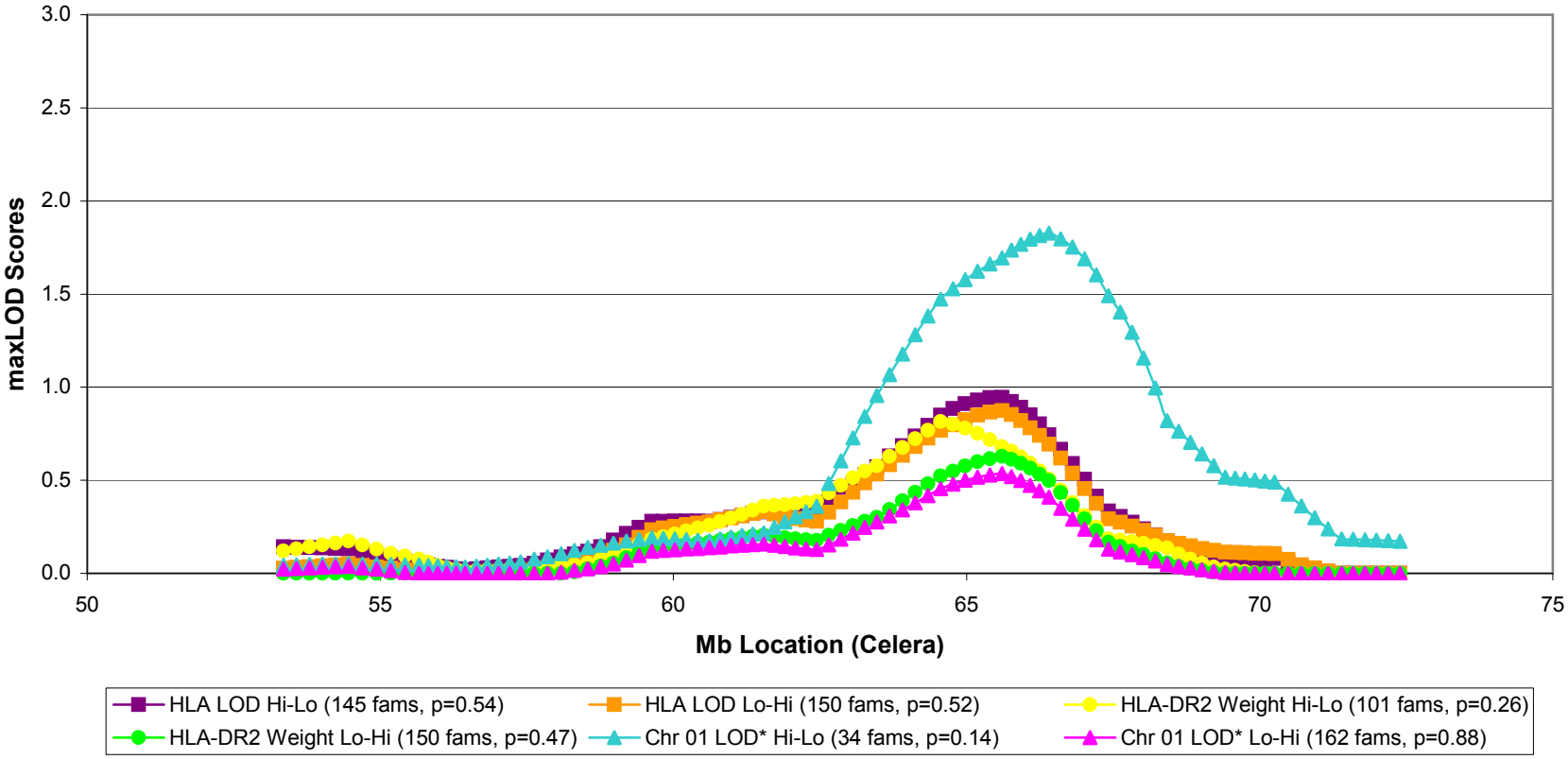


Figure 15. Chromosome 16q Multipoint Linkage Plot – Ordered Subset Analysis

Table 6. Chromosome 18p Analysis

						MLOD			PDT		
SNP ID (Celera)	SNP ID (dbSNP)	Mb (Celera)	Mb (dbSNP)	Gene	MAF	All	HLA All	HLA None	PDTSum	PDTAve	PDTGeno
hCV1367907	rs585578	0.079384	194232	USP14/THOC1	0.34	0.25	0.00	0.99	0.129	0.778	0.261
hCV3013161	rs1608446	0.790552	905946	ADCYAP1	0.35	0.00	0.00	0.74	0.829	0.268	0.359
hCV3235262	rs313021	1.758686	1873794	N/A	0.36	0.00	0.00	0.44	0.764	0.442	0.747
GATA166D05		1.998573			8 alleles	0.20	0.11	0.67	0.845	0.570	0.562
hCV3011859	rs598866	2.760875	2875437	EMILIN-2	0.45	0.00	0.00	0.14	0.117	0.146	0.269
hCV8121217	rs11664127	3.866213	3972406	DLGAP1	0.43	0.09	0.12	0.06	0.863	0.458	0.856
hCV11205167	rs9963665	5.117272	5222843	MGC17515	0.47	0.00	0.00	0.75	0.501	0.460	0.207
hCV1367564	rs736632	5.859604	5963728	FLJ35936	0.50	0.30	0.00	**1.40**	0.294	0.339	0.630
D18S967		6.481905			5 alleles	0.14	0.00	0.28	0.238	0.183	0.701
hCV3086198	rs2089760	7.004187	7108955	LAMA1	0.49	0.00	0.00	0.75	0.890	0.746	0.927
hCV1464431	N/A	7.897796	8003798	PTPRM	0.42	0.00	0.00	0.89	0.561	0.812	0.873
D18S843		8.498119			6 alleles	0.38	0.00	**2.84**	0.593	0.358	0.815
hCV7495855	rs977581	9.003318	9109035	NDUFV2	0.39	0.20	0.20	0.47	0.237	0.465	0.157
hCV612923	rs29062	9.848813	9955338	VAPA	0.45	0.74	0.07	0.49	0.793	0.690	0.784
hCV1637378	rs1026390	11.365124	11473169	N/A	0.41	0.89	0.49	0.67	**0.014**	**0.045**	**0.027**
hCV460143	rs4797665	12.159784	12269553	CIDEA	0.50	0.07	0.00	0.07	0.241	0.504	0.265
hCV8116885	rs872906	13.259732	13367699	C18orf1	0.40	0.00	0.00	0.69	0.813	0.875	0.978
hCV3047081	rs9956386	13.981109	14088926	LOC162655	0.50	0.00	0.00	0.00	0.106	0.364	0.064
hCV15827884	rs2127958	15.459255	16907608	ROCK1	0.43	0.49	0.00	0.74	0.562	0.962	0.608
hCV16100206	rs2847129	16.040816	17488816	ABHD3	0.34	0.93	0.00	**1.62**	0.506	0.792	0.519
D18S869		16.891612			9 alleles	0.21	0.01	0.34	0.659	0.602	0.932
hCV194412	rs11082010	16.986539	18426403	hCG1643126	0.50	0.36	0.15	0.25	0.843	0.969	0.962
hCV3204015	rs6507720	17.935283	19373289	MIC1/NPC1	0.48	0.12	0.00	0.20	0.077	0.117	0.188
hCV2019690	rs600958	18.906057	20345021	hCG1643857	0.49	0.53	0.00	0.49	0.588	0.293	0.557
hCV3253578	rs273756	20.007863	21447335	hCG1656949	0.49	0.28	0.51	0.05	0.772	0.906	0.968

hCV3004492	rs2438414	20.999322	22442676	hCG38480	0.45	0.11	0.00	0.60	0.459	0.363	0.680
hCV7492337	rs1467233	21.837767	23281358	N/A	0.29	0.06	0.00	0.01	0.615	0.895	0.878
hCV3117911	rs1354417	23.044179	24486928	hCG1641502	0.48	0.29	0.09	0.32	0.443	0.539	0.065
hCV11732665	rs12605279	23.778441	25226212	N/A	0.48	0.18	0.09	0.20	0.126	0.179	0.207

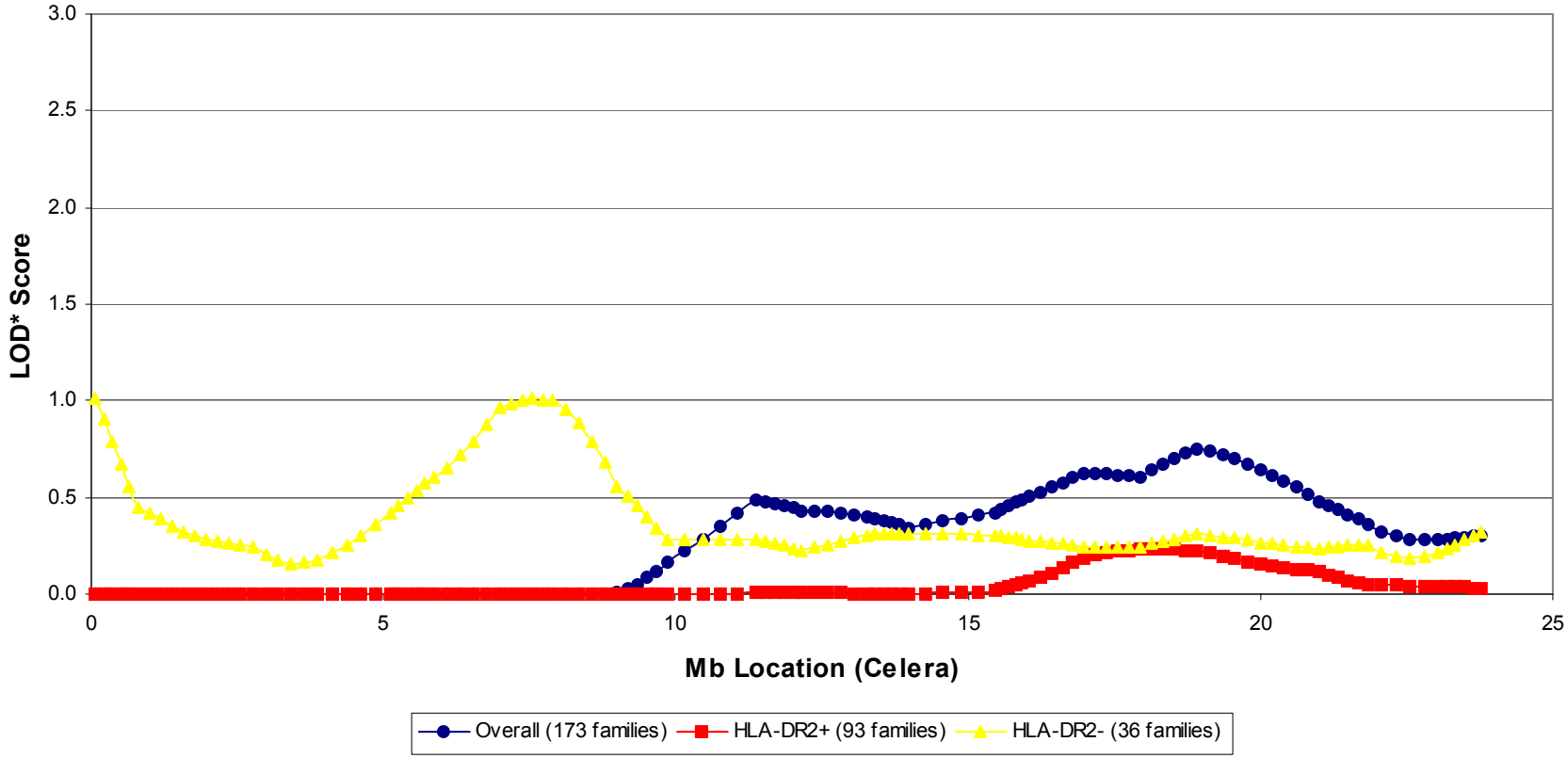


Figure 16. Chromosome 18p Multipoint Linkage Plot

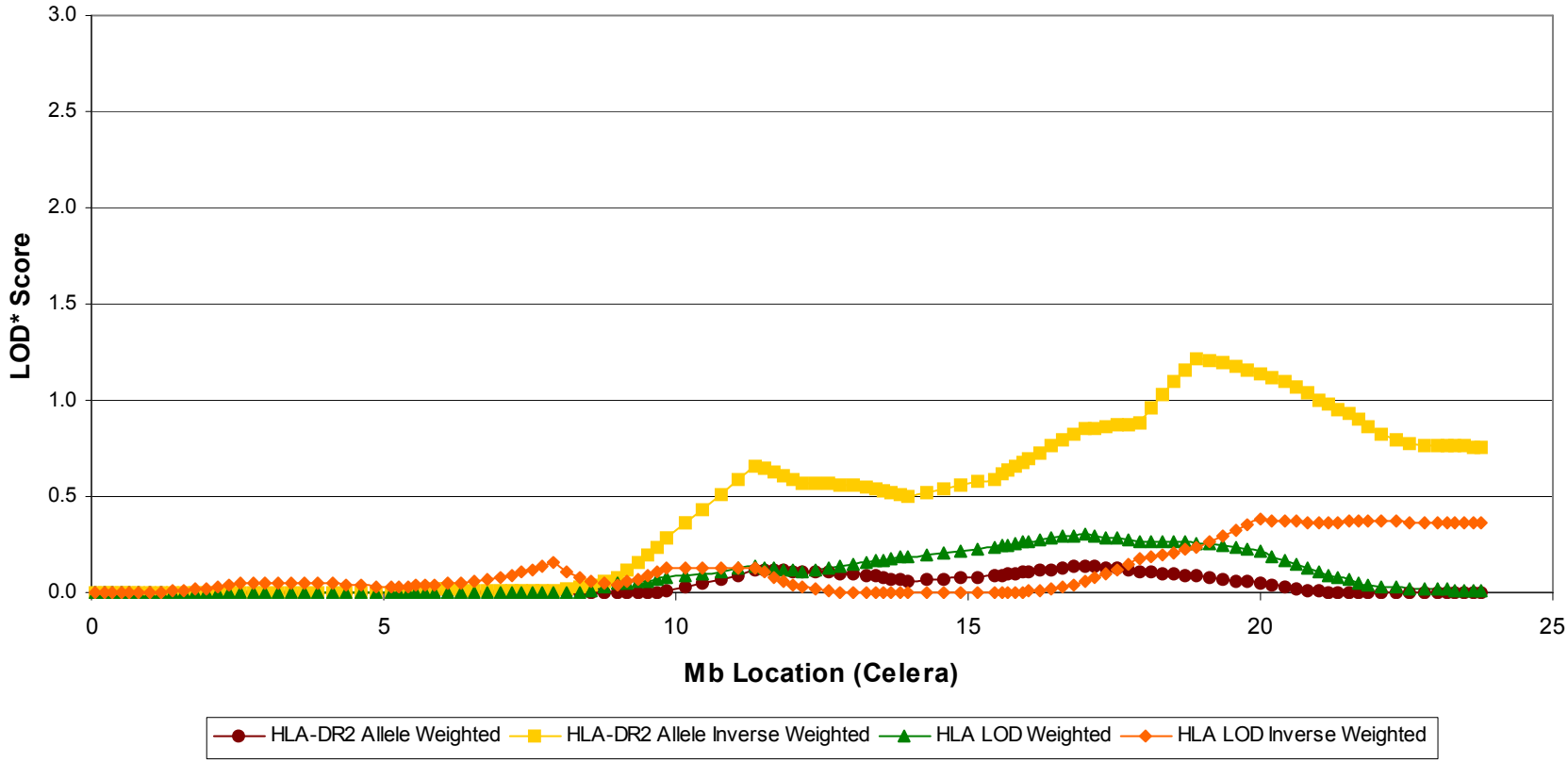


Figure 17. Chromosome 18p Multipoint Linkage Plot – HLA Conditional Analysis

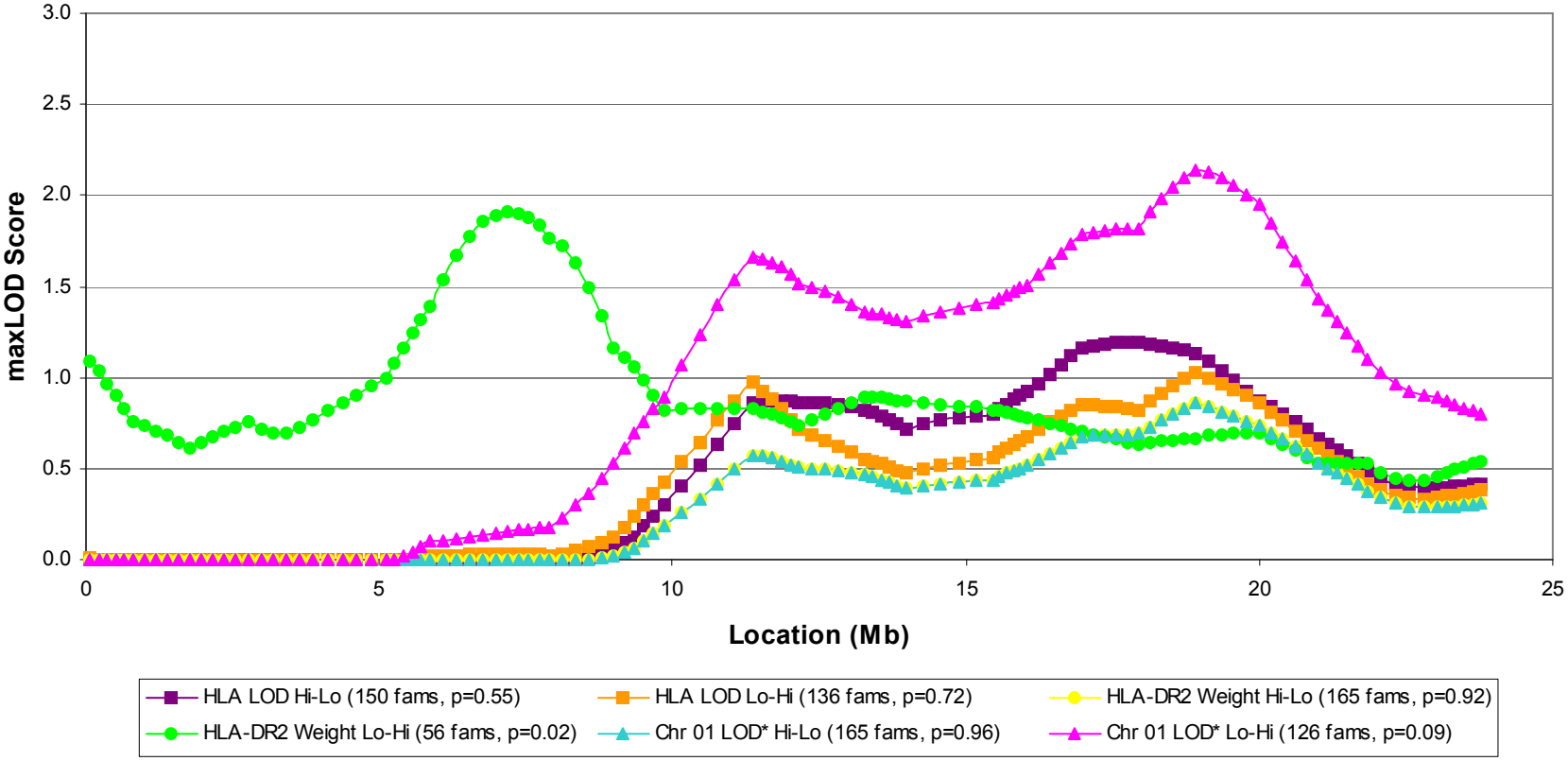


Figure 18. Chromosome 18p Multipoint Linkage Plot – Ordered Subset Analysis

Table 7. Chromosome 19q Analysis

						MLOD			PDT		
SNP ID (Celera)	**SNP ID (dbSNP)**	**Mb (Celera)**	**Mb (dbSNP)**	**Gene**	**MAF**	**All**	**HLA All**	**HLA None**	**PDTSum**	**PDTAve**	**PDTGeno**
hCV7610932	rs1035478	30.009929	38008163	SLC7A9	0.46	0.01	0.21	0.00	0.727	0.734	0.492
hCV2592102	rs285694	30.954616	38952829	CHST8	0.46	0.09	0.22	0.03	0.619	0.413	0.887
D19S587		31.922494			7 alleles	0.00	0.07	0.00	0.425	0.271	0.235
hCV25473593	rs8100085	32.236717	40214959	SCN1B/HPN	0.35	0.00	0.08	0.00	0.076	**0.042**	0.069
hCV3111700	rs120960	32.99024	40968607	SNX26	0.43	0.84	0.45	0.25	0.114	0.378	0.064
hCV1277048	rs826285	34.163327	42070872	N/A	0.49	0.20	0.00	0.25	0.772	0.274	**0.049**
hCV1135244	rs953370	35.183266	43074168	NYD-SP11	0.38	**1.35**	0.55	0.43	0.650	0.131	0.900
hCV11507800	rs3745859	35.998759	43888585	ACTN4/M9	0.45	0.22	0.59	0.06	0.927	0.974	0.167
hCV2275630	rs17608	37.027214	44917485	CLC	0.34	0.65	0.17	0.42	0.627	0.657	0.885
hCV11465156	rs1870087	38.056553	45947971	hCG20793	0.48	0.08	0.22	0.14	0.324	0.814	0.527
hCV8597405	N/A	39.065597	46957728	CEACAM6	0.43	0.08	0.04	0.11	0.798	0.628	0.693
D19S211		40.184511			12 alleles	**1.79**	0.70	0.97	0.520	0.558	0.269
hCV11975183	rs2024096	40.467154	48358401	PSG5	0.43	0.28	0.30	0.21	0.977	0.307	0.091
hCV2652664	rs417699	41.221098	49109415	ZNF45	0.50	0.21	0.47	0.05	0.074	0.282	0.069
hCV3084818	rs760136	42.208180	50095698	APOE	0.43	0.00	0.27	0.09	0.680	0.720	0.797
APOE	N/A	42.216263	N/A	APOE	3 alleles	0.47	0.47	0.22	0.836	0.835	0.792
hCV2884324	rs7256192	43.143522	51030332	SPK	0.35	0.23	0.18	0.75	0.465	0.808	0.486
hCV3168164	rs10425791	44.019503	51907814	STRN4/PKD2	0.34	0.59	0.50	0.00	0.617	0.872	0.849
hCV1996744	rs3786780	45.059369	52945479	EHD2/GLTSCR2	0.48	**1.46**	0.58	0.55	0.727	0.417	0.794
hCV2405293	rs602662	46.073085	53898797	FUT2	0.50	0.62	0.99	0.16	0.495	0.694	0.384
hCV1844598	rs6509443	47.114353	54936708	TSKS/CPT1C	0.46	0.00	0.00	0.01	0.273	0.175	0.568
hCV1655600	rs1673028	48.003223	55644865	MYBPC2	0.38	0.68	0.36	0.90	0.440	0.666	0.720
hCV3057052	rs11668530	49.052525	56693297	SIGLECL1	0.47	0.05	0.10	0.00	0.360	0.641	0.366
D19S402		49.220919			18 alleles	0.11	0.16	0.00	0.100	0.042	0.629
hCV3098516	rs619872	50.201641	57844212	ZNF83	0.48	0.62	0.64	0.08	0.689	0.773	0.931
hCV1997401	rs3848580	51.329927	58980658	N/A	0.49	**2.01**	0.94	0.19	0.899	0.434	0.295
hCV2996831	rs270790	52.279937	59931035	KIR3DL7	0.47	0.93	0.49	0.08	0.066	0.570	0.287
hCV8879227	rs7478	53.038441	60690323	KLP1/HSPC189	0.40	0.71	**1.42**	0.29	0.764	0.834	0.741

hCV2658107	rs3760849	53.996589	61645397	FLJ14011	0.47	**1.79**	0.85	**1.23**	0.980	0.996	0.994
hCV11701198	rs7250521	54.913856	62564959	ZNF304/ZNF547	0.49	0.11	0.60	0.04	0.248	0.060	0.496
hCV1116794	rs1465789	55.990054	63637868	ZNF132	0.44	0.11	0.36	0.00	0.680	0.113	0.912

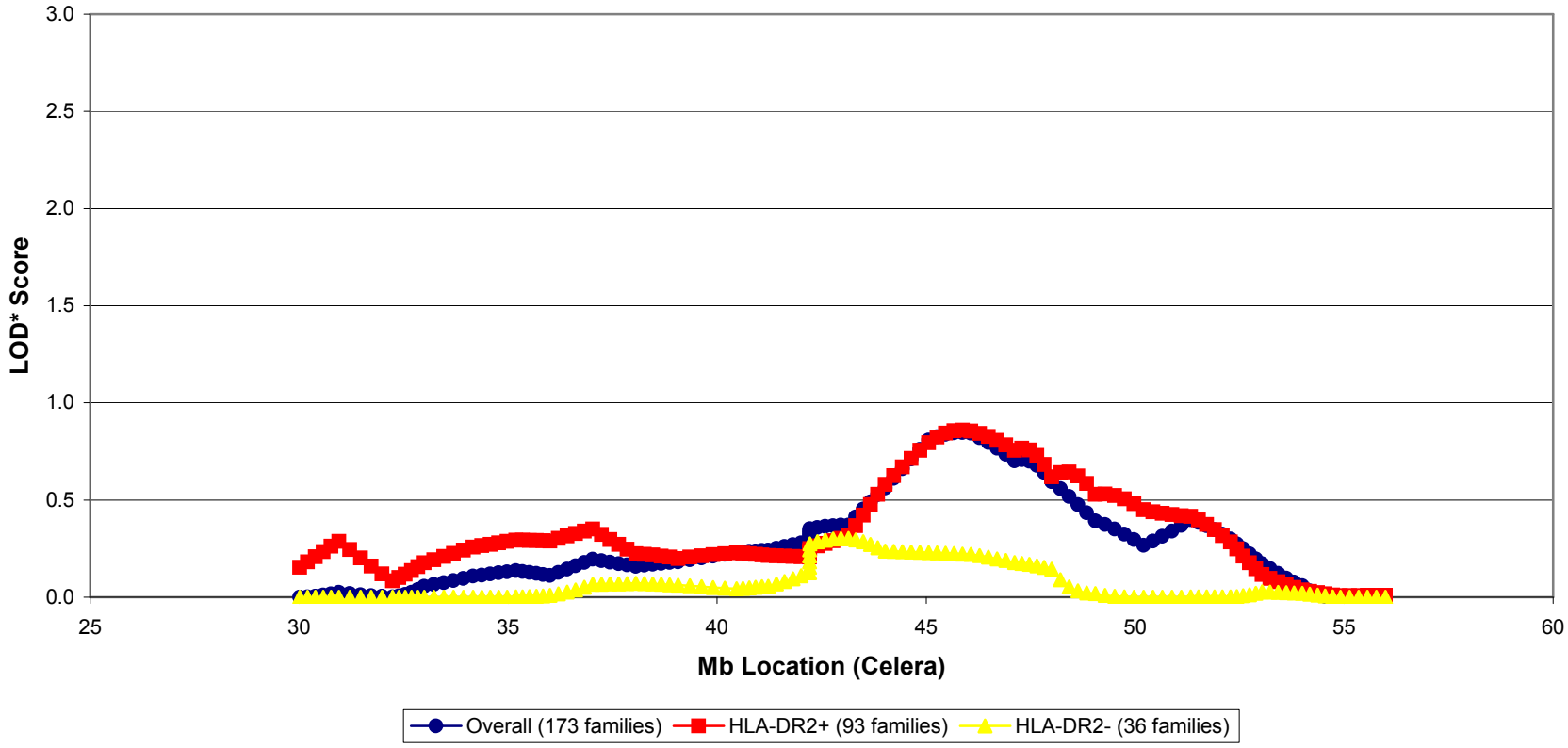


Figure 19. Chromosome 19q Multipoint Linkage Plot

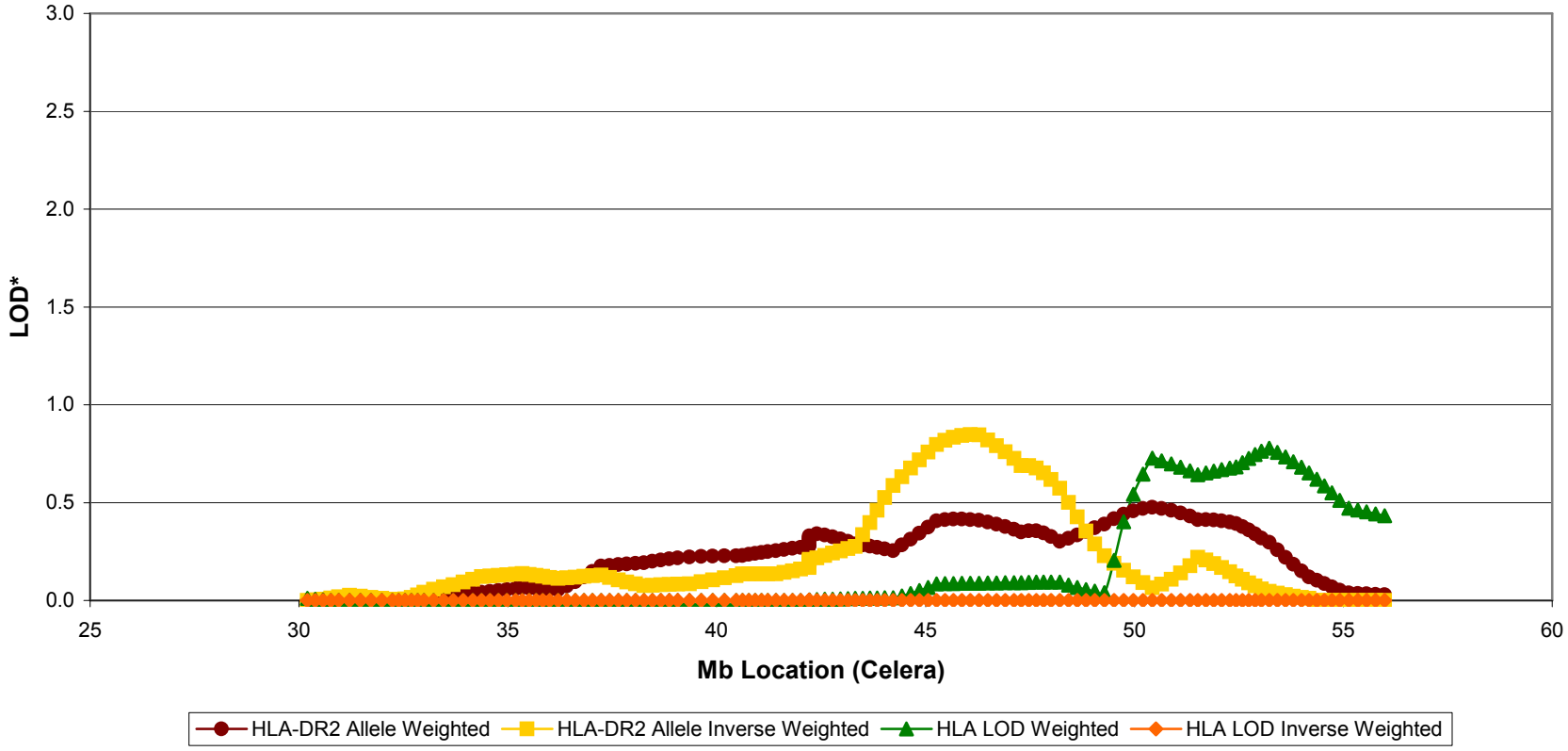


Figure 20. Chromosome 19q Multipoint Linkage Plot – HLA Conditional Analysis

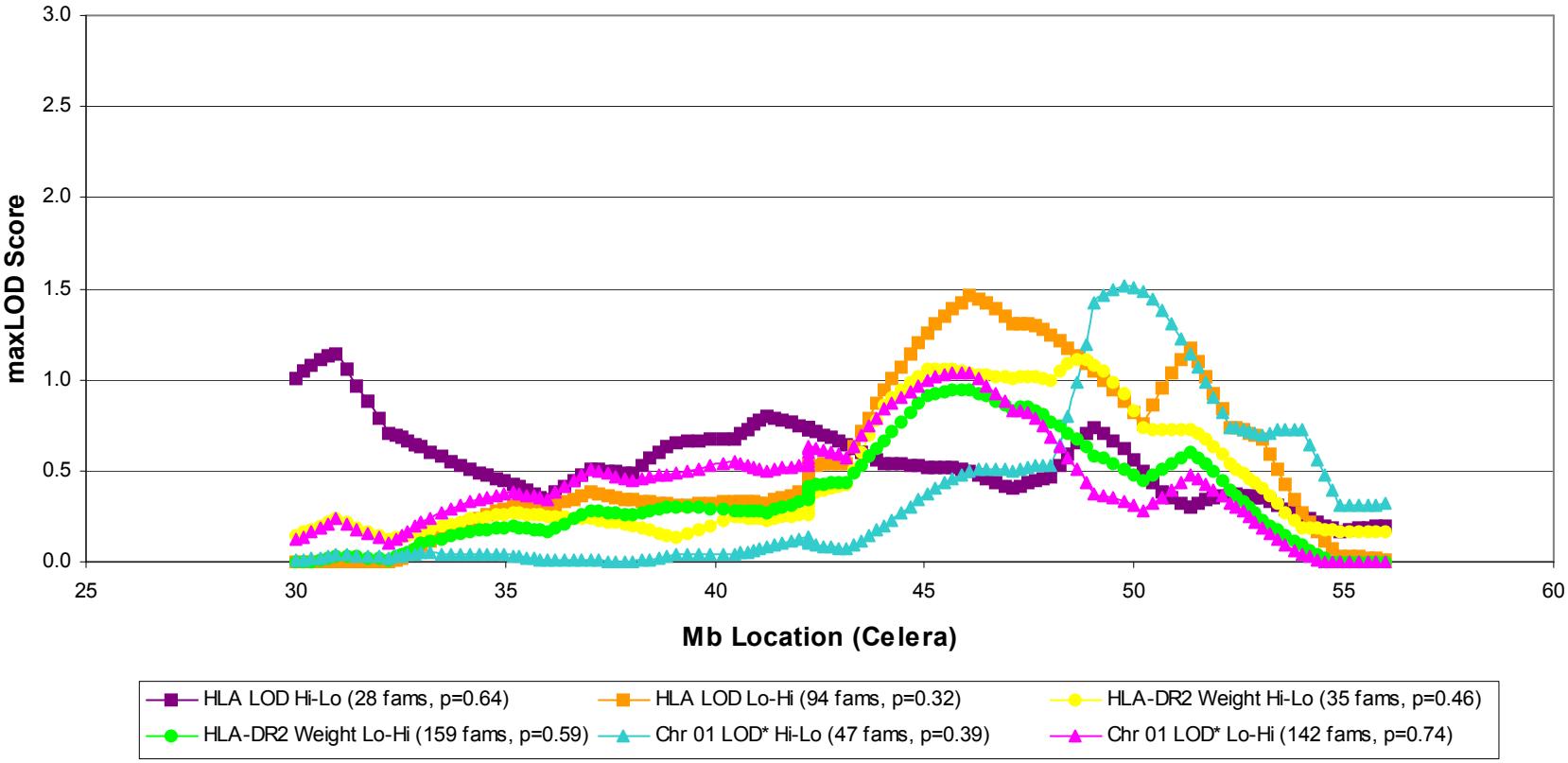


Figure 21. Chromosome 19q Multipoint Linkage Plot – Ordered Subset Analysis

APPENDIX D

Table 1. Illumina SNPs – Chromosome 1q44

Public ID	NCBI Mb Location*	Mb Interval	MAF
rs12123449	234.248599	0.015288	0.155
rs12129023	234.263887	0.005178	0.199
rs9428368	234.269065	0.020787	0.204
rs6428989	234.289852	0.013481	0.043
rs7551672	234.303333	0.015427	0.082
rs1933129	234.318760	0.018620	0.062
rs10737813	234.337380	0.023289	0.411
rs2038889	234.360669	0.000258	0.354
rs1337799	234.360927	0.027230	0.011
rs792553	234.388157	0.010004	0.005
rs1765887	234.398161	0.007570	0.130
rs12401834	234.405731	0.010575	0.381
rs618083	234.416306	0.002998	0.490
rs2152884	234.419304	0.015122	0.230
rs1362841	234.434426	0.029776	0.358
rs961121	234.464202	0.010572	0.048
rs2275287	234.474774	0.026137	0.297
rs7532774	234.500911	0.009542	0.424
rs10737814	234.510453	0.027948	0.355
rs6429005	234.538401	0.013662	0.430
rs2050656	234.552063	0.009242	0.148
rs4336842	234.561305	0.019197	0.211
rs10925388	234.580502	0.001262	0.166
rs10925391	234.581764	0.010761	0.247
rs10754602	234.592525	0.014849	0.442
rs4531285	234.607374	0.000164	0.314
rs12138118	234.607538	0.008906	0.005
rs4465196	234.616444	0.000098	0.309
rs4659791	234.616542	0.000174	0.309
rs4659792	234.616716	0.042190	0.309
rs3765097	234.658906	0.006173	0.434
rs2127153	234.665079	0.001929	0.499
rs1478913	234.667008	0.005695	0.500
rs12094480	234.672703	0.009513	0.011
rs10495396	234.682216	0.017571	0.492
rs2010032	234.699787	0.004473	0.460
rs2805422	234.704260	0.007029	0.335
rs2618702	234.711289	0.003731	0.308
rs3766841	234.715020	0.001161	0.201

rs4336838	234.716181	0.000223	0.002
rs2618651	234.716404	0.017935	0.304
rs2805446	234.734339	0.018607	0.199
rs2253273	234.752946	0.013930	0.038
rs722581	234.766876	0.021286	0.339
rs7554494	234.788162	0.002559	0.438
rs1478914	234.790721	0.001605	0.409
rs1564272	234.792326	0.001381	0.253
rs2779427	234.793707	0.000806	0.005
rs2805409	234.794513	0.002038	0.433
rs2618714	234.796551	0.001625	0.434
rs1478915	234.798176	0.000343	0.433
rs12136895	234.798519	0.000151	0.256
rs1478916	234.798670	0.000092	0.402
rs3766869	234.798762	0.009140	0.011
rs12137565	234.807902	0.006674	0.386
rs1717783	234.814576	0.004657	0.355
rs3766871	234.819233	0.000825	0.025
rs596502	234.820058	0.035874	0.167
rs684923	234.855932	0.003927	0.435
rs3766881	234.859859	0.023491	0.243
rs12121446	234.883350	0.009040	0.059
rs12074235	234.892390	0.009772	0.065
rs12079834	234.902162	0.010898	0.276
rs12127746	234.913060	0.034696	0.090
rs2819771	234.947756	0.000972	0.392
rs12057693	234.948728	0.008569	0.319
rs10495401	234.957297	0.002358	0.430
rs6670609	234.959655	0.008134	0.024
rs12080621	234.967789	0.001581	0.001
rs10802632	234.969370	0.000631	0.355
rs12404009	234.970001	0.018112	0.055
rs790889	234.988113	0.010197	0.334
rs790901	234.998310	0.000148	0.285
rs790900	234.998458	0.010628	0.291
rs2794828	235.009086	0.004752	0.225
rs790882	235.013838	0.022594	0.219
rs4659819	235.036432	0.000388	0.479
rs12025731	235.036820	0.000166	0.482
rs790897	235.036986	0.000240	0.005
rs12725752	235.037226	0.001010	0.007
rs12594	235.038236	0.017883	0.322
rs1890672	235.056119	0.015957	0.287
rs12092452	235.072076	0.015676	0.009
rs2487789	235.087752	0.020633	0.373

rs559849	235.108385	0.036913	0.235
rs2499602	235.145298	0.000092	0.434
rs2451615	235.145390	0.001825	0.433
rs960292	235.147215	0.000063	0.003
rs960291	235.147278	0.002353	0.003
rs902333	235.149631	0.000853	0.257
rs2499595	235.150484	0.002158	0.434
rs2451609	235.152642	0.000332	0.441
rs2499589	235.152974	0.000510	0.441
rs902332	235.153484	0.000077	0.443
rs902331	235.153561	0.001013	0.438
rs1565426	235.154574	0.000337	0.441
rs2499586	235.154911	0.000481	0.441
rs2248471	235.155392	0.007367	0.440
rs6663760	235.162759	0.003609	0.484
rs747520	235.166368	0.000570	0.038
rs902329	235.166938	0.036076	0.485
rs1017981	235.203014	0.025645	0.477
rs6672409	235.228659	0.024936	0.475
rs12087117	235.253595	0.017424	0.008
rs1338920	235.271019	0.004227	0.264
rs2487084	235.275246	0.025440	0.228
rs1417275	235.300686	0.010502	0.059
rs1796942	235.311188	0.016991	0.061
rs10495405	235.328179	0.020877	0.057
rs1416385	235.349056	0.026408	0.178
rs6413954	235.375464	0.028795	0.178
rs922694	235.404259	0.023532	0.467
rs10737820	235.427791	0.026128	0.364
rs7542189	235.453919	0.026538	0.458
rs10495407	235.480457	0.026955	0.352
rs12077136	235.507412	0.046080	0.014
rs7538546	235.553492	0.030964	0.057
rs10802681	235.584456	0.015777	0.398
rs6698025	235.600233	0.026334	0.313
rs7547119	235.626567	0.012794	0.181
rs1342079	235.639361	0.028715	0.203
rs7550966	235.668076	0.007026	0.238
rs12029859	235.675102	0.003772	0.172
rs10754638	235.678874	0.006375	0.323
rs1054888	235.685249	0.000190	0.322
rs1557131	235.685439	0.003813	0.328
rs2153613	235.689252	0.002846	0.287
rs10495411	235.692098	0.021207	0.236
rs2893644	235.713305	0.023489	0.270

rs10495417	235.736794	0.023513	0.095
rs526612	235.760307	0.039373	0.092
rs586565	235.799680	0.018731	0.260
rs632407	235.818411	0.000266	0.048
rs574819	235.818677	0.029249	0.228
rs12078080	235.847926	0.031089	0.005
rs2174076	235.879015	0.015580	0.479
rs2653912	235.894595	0.029094	0.203
rs4659885	235.923689	0.011706	0.376
rs10158346	235.935395	0.036815	0.402
rs2841417	235.972210	0.031538	0.317
rs2689163	236.003748	0.000069	0.461
rs10925763	236.003817	0.024552	0.177
rs12406780	236.028369	0.025955	0.096
rs2689138	236.054324	0.033304	0.151
rs9428749	236.087628	0.010558	0.145
rs9428417	236.098186	0.022665	0.146
rs1915251	236.120851	0.031388	0.332
rs1915258	236.152239	0.004818	0.007
rs9428788	236.157057	0.013787	0.143
rs2139665	236.170844	0.045584	0.325
rs1915245	236.216428	0.046472	0.481
rs1912230	236.262900	0.026992	0.074
rs6682504	236.289892	0.005997	0.288
rs2353429	236.295889	0.020839	0.236
rs10495440	236.316728	0.019517	0.194
rs6685861	236.336245	0.026161	0.345
rs7526587	236.362406	0.024535	0.329
rs967290	236.386941	0.031835	0.304
rs7521497	236.418776	0.014143	0.315
rs1339737	236.432919	0.024999	0.332
rs10925857	236.457918	0.012074	0.158
rs12405815	236.469992	0.002402	0.026
rs2820033	236.472394	0.028922	0.158
rs4579742	236.501316	0.018836	0.154
rs12042304	236.520152	0.012831	0.138
rs6661899	236.532983	0.007081	0.323
rs9428828	236.540064	0.055605	0.254
rs6429136	236.595669	0.001073	0.114
rs6675386	236.596742	0.031919	0.114
rs7548324	236.628661	0.024676	0.303
rs6681143	236.653337	0.019867	0.016
rs10754671	236.673204	0.024732	0.469
rs10399730	236.697936	0.022079	0.007
rs1782356	236.720015	0.005149	0.438

rs1782351	236.725164	0.006829	0.443
rs1218671	236.731993	0.005537	0.378
rs6676678	236.737530	0.002743	0.409
rs9287229	236.740273	0.009860	0.413
rs1416379	236.750133	0.004675	0.083
rs10925915	236.754808	0.012588	0.414
rs1984207	236.767396	0.027715	0.414
rs10754674	236.795111	0.003230	0.075
rs997538	236.798341	0.026023	0.080
rs6697471	236.824364	0.016709	0.234
rs12096150	236.841073	0.026735	0.051
rs2355232	236.867808	0.028318	0.358
rs7523711	236.896126	0.000311	0.017
rs12028626	236.896437	0.003437	0.062
rs6429153	236.899874	0.023153	0.489
rs12047255	236.923027	0.001377	0.062
rs7511970	236.924404	0.000361	0.451
rs6694220	236.924765	0.007386	0.462
rs10429918	236.932151	0.011512	0.062
rs10925964	236.943663	0.002565	0.450
rs1019881	236.946228	0.004863	0.469
rs10802802	236.951091	0.018975	0.447
rs654209	236.970066	0.010979	0.253
rs485412	236.981045	0.022604	0.262
rs569474	237.003649	0.007342	0.468
rs552634	237.010991	0.000724	0.222
rs10925980	237.011715	0.003440	0.261
rs6690809	237.015155	0.021283	0.486
rs576386	237.036438	0.003336	0.422
rs7548522	237.039774	0.047283	0.062
rs536477	237.087057	0.009565	0.405
rs10495448	237.096622	0.000085	0.302
rs12068071	237.096707	0.004308	0.019
rs6684778	237.101015	0.001769	0.091
rs10802816	237.102784	0.005625	0.063
rs7520974	237.108409	0.003343	0.430
rs3738435	237.111752	0.000341	0.169
rs2067481	237.112093	0.024031	0.018
rs7531091	237.136124	0.013568	0.283
rs12563039	237.149692	0.027359	0.487
rs4233479	237.177051	0.000069	0.195
rs12133285	237.177120	0.000280	0.336
rs4659558	237.177400	0.004324	0.086
rs1381525	237.181724	0.021146	0.492
rs7512434	237.202870	0.016548	0.318

rs10436944	237.219418	0.005533	0.433
rs2066299	237.224951	0.018002	0.463
rs9887796	237.242953	0.019587	0.265
rs12069776	237.262540	0.000228	0.174
rs12095464	237.262768	0.001322	0.172
rs12094054	237.264090	0.017084	0.175
rs9287230	237.281174	0.046428	0.154
rs3738434	237.327602	0.000263	0.038
rs3738433	237.327865	0.002536	0.076
rs7524136	237.330401	0.014070	0.070
rs7537876	237.344471	0.002545	0.444
rs12073759	237.347016	0.000250	0.002
rs12069275	237.347266	0.000337	0.005
rs12071494	237.347603	0.000935	0.008
rs2355833	237.348538	0.000433	0.326
rs12087179	237.348971	0.000070	0.002
rs1539200	237.349041	0.000069	0.345
rs1471089	237.349110	0.000239	0.196
rs12127949	237.349349	0.007768	0.347
rs10495459	237.357117	0.025300	0.449
rs3765588	237.382417	0.021973	0.190
rs1456661	237.404390	0.001093	0.304
rs10802846	237.405483	0.006295	0.366
rs10926166	237.411778	0.012155	0.387
rs6429189	237.423933	0.024199	0.239
rs1456654	237.448132	0.000180	0.370
rs1456655	237.448312	0.001409	0.368
rs1456658	237.449721	0.012186	0.367
rs4659951	237.461907	0.000679	0.381
rs10926182	237.462586	0.000521	0.346
rs10926184	237.463107	0.000405	0.345
rs10802850	237.463512	0.021420	0.346
rs7554349	237.484932	0.000386	0.147
rs7531591	237.485318	0.000136	0.122
rs2045349	237.485454	0.029399	0.267
rs1379074	237.514853	0.004681	0.165
rs4433380	237.519534	0.009162	0.096
rs10495463	237.528696	0.004016	0.469
rs10157903	237.532712	0.000851	0.366
rs6677726	237.533563	0.000320	0.085
rs3795677	237.533883	0.000968	0.281
rs3795680	237.534851	0.003563	0.288
rs10157874	237.538414	0.007861	0.478
rs4659963	237.546275	0.014842	0.413
rs1932589	237.561117	0.011201	0.484

rs12089062	237.572318	0.034699	0.243
rs4659971	237.607017	0.005042	0.351
rs12028974	237.612059	0.006231	0.169
rs9661317	237.618290	0.009554	0.007
rs1414660	237.627844	0.012404	0.190
rs10926254	237.640248	0.002170	0.336
rs2356386	237.642418	0.000363	0.093
rs2356387	237.642781	0.000242	0.263
rs9728292	237.643023	0.000127	0.213
rs9728305	237.643150	0.000167	0.289
rs12143600	237.643317	0.005728	0.080
rs12065307	237.649045	0.002358	0.144
rs945512	237.651403	0.002603	0.247
rs12064827	237.654006	0.000850	0.187
rs10802871	237.654856	0.019840	0.297
rs12064536	237.674696	0.000125	0.001
rs9727156	237.674821	0.003899	0.392
rs882869	237.678720	0.013262	0.336
rs1336153	237.691982	0.004182	0.311
rs3748535	237.696164	0.001782	0.399
rs11806449	237.697946	0.004191	0.237
rs12133916	237.702137	0.001171	0.463
rs12569318	237.703308	0.004843	0.034
rs10465630	237.708151	0.008197	0.246
rs7540147	237.716348	0.008290	0.368
rs10802883	237.724638	0.007418	0.350
rs6668956	237.732056	0.003095	0.272
rs3748538	237.735151	0.002212	0.346
rs2185283	237.737363	0.009893	0.346
rs10495471	237.747256	0.001844	0.127
rs6429210	237.749100	0.003969	0.425
rs11588607	237.753069	0.001305	0.332
rs11582912	237.754374	0.003236	0.330
rs7411138	237.757610	0.010297	0.098
rs7528086	237.767907	0.014188	0.023
rs4453026	237.782095	0.014298	0.266
rs10495472	237.796393	0.001605	0.103
rs1934342	237.797998	0.013939	0.276
rs6686630	237.811937	0.001372	0.447
rs10495473	237.813309	0.001123	0.288
rs7544440	237.814432	0.001820	0.438
rs3748533	237.816252	0.025013	0.189
rs2185288	237.841265	0.014235	0.114
rs6429216	237.855500	0.026697	0.460
rs6702786	237.882197	0.010129	0.133

rs10926334	237.892326	0.007457	0.078
rs7546208	237.899783	0.004062	0.432
rs10754715	237.903845	0.010191	0.277
rs12097614	237.914036	0.003391	0.001
rs10495474	237.917427	0.001984	0.338
rs6676703	237.919411	0.001438	0.014
rs6683045	237.920849	0.006946	0.299
rs7518789	237.927795	0.042389	0.056
rs11586149	237.970184	0.002669	0.166
rs717791	237.972853	0.001741	0.268
rs6677808	237.974594	0.000627	0.030
rs6695843	237.975221	0.000765	0.031
rs4568818	237.975986	0.029396	0.390
rs4603122	238.005382	0.010980	0.313
rs2275742	238.016362	0.004590	0.347
rs3738068	238.020952	0.003374	0.186
rs4659581	238.024326	0.003115	0.186
rs10926365	238.027441	0.005279	0.388
rs12410838	238.032720	0.013399	0.087
rs1977840	238.046119	0.005934	0.221
rs785976	238.052053	0.034708	0.148
rs377116	238.086761	0.022072	0.453
rs261806	238.108833	0.020024	0.453
rs422256	238.128857	0.002445	0.438
rs195778	238.131302	0.005423	0.381
rs261861	238.136725	0.023765	0.478
rs2502436	238.160490	0.015595	0.124
rs3893178	238.176085	0.014024	0.457
rs10802917	238.190109	0.012815	0.482
rs670659	238.202924	0.009202	0.340
rs628208	238.212126	0.029018	0.429
rs538423	238.241144	0.010260	0.389
rs4660027	238.251404	0.052631	0.348
rs12083715	238.304035	0.023559	0.002
rs4634894	238.327594	0.007023	0.409
rs7548582	238.334617	0.000320	0.001
rs7550998	238.334937	0.026719	0.001
rs10465612	238.361656	0.005919	0.407
rs6665888	238.367575	0.015865	0.384
rs4391653	238.383440	0.025430	0.256
rs2177113	238.408870	0.012178	0.080
rs4233483	238.421048	0.018699	0.327
rs2815848	238.439747	0.035870	0.410
rs10926448	238.475617	0.003615	0.189
rs10926450	238.479232	0.001589	0.469

rs4659596	238.480821	0.014835	0.469
rs6697953	238.495656	0.026137	0.264
rs12121748	238.521793	0.003126	0.016
rs1341468	238.524919	0.007031	0.407
rs4659598	238.531950	0.022162	0.225
rs4660068	238.554112	0.006135	0.232
rs4659599	238.560247	0.001090	0.015
rs1891130	238.561337	0.005334	0.025
rs1341463	238.566671	0.000954	0.235
rs10926467	238.567625	0.023874	0.125
rs1557078	238.591499	0.010700	0.477
rs10802960	238.602199	0.012497	0.407
rs9428505	238.614696	0.022546	0.138
rs3014554	238.637242	0.015798	0.284
rs1954202	238.653040	0.022629	0.239
rs6696900	238.675669	0.003016	0.001
rs10926494	238.678685	0.027974	0.325
rs2185626	238.706659	0.011525	0.413
rs2275162	238.718184	0.008151	0.027
rs4660100	238.726335	0.010606	0.269
rs3819976	238.736941	0.010240	0.025
rs11805494	238.747181	0.006102	0.404
rs12410855	238.753283	0.017920	0.390
rs2275163	238.771203	0.022213	0.305
rs2273711	238.793416	0.003081	0.010
rs1053230	238.796497	0.000219	0.237
rs1053221	238.796716	0.001982	0.138
rs3765809	238.798698	0.001082	0.170
rs1053183	238.799780	0.009077	0.199
rs2273712	238.808857	0.013074	0.010
rs3753221	238.821931	0.017849	0.240
rs3737604	238.839780	0.002766	0.318
rs581510	238.842546	0.020785	0.282
rs7416113	238.863331	0.014795	0.241
rs6661311	238.878126	0.012070	0.342
rs7527828	238.890196	0.016172	0.386
rs12117802	238.906368	0.042472	0.322
rs7418599	238.948840	0.000126	0.382
rs10802990	238.948966	0.004857	0.459
rs6669629	238.953823	0.000634	0.208
rs6429302	238.954457	0.012570	0.161
rs7551270	238.967027	0.003664	0.208
rs12095445	238.970691	0.017224	0.159
rs953597	238.987915	0.000444	0.315
rs2051064	238.988359	0.004119	0.484

rs12074374	238.992478	0.001972	0.169
rs6429314	238.994450	0.007368	0.378
rs1342866	239.001818	0.027591	0.164
rs10489351	239.029409	0.023773	0.200
rs1635517	239.053182	0.000352	0.410
rs4149852	239.053534	0.000169	0.079
rs1776176	239.053703	0.000106	0.492
rs2073490	239.053809	0.022722	0.201
rs4149963	239.076531	0.006919	0.072
rs1047840	239.083450	0.000063	0.383
rs12122770	239.083513	0.000181	0.006
rs1776148	239.083694	0.002730	0.356
rs1635498	239.086424	0.007346	0.009
rs4150018	239.093770	0.023969	0.436
rs5009401	239.117739	0.002782	0.459
rs7512631	239.120521	0.000126	0.001
rs7555402	239.120647	0.006102	0.462
rs7517808	239.126749	0.012103	0.219
rs3844254	239.138852	0.031222	0.066
rs10158939	239.170074	0.040752	0.326
rs12124966	239.210826	0.012015	0.213
rs6663013	239.222841	0.008295	0.468
rs6672510	239.231136	0.037673	0.343
rs316839	239.268809	0.000764	0.051
rs12134852	239.269573	0.006942	0.047
rs1393299	239.276515	0.002161	0.433
rs316894	239.278676	0.023684	0.046
rs316822	239.302360	0.028067	0.439
rs12408396	239.330427	0.023208	0.259
rs421625	239.353635	0.021824	0.178
rs12405001	239.375459	0.023121	0.009
rs2841902	239.398580	0.018338	0.122
rs10803034	239.416918	0.003348	0.101
rs4589096	239.420266	0.003713	0.142
rs3863747	239.423979	0.054399	0.240
rs1439523	239.478378	0.000830	0.030
rs977112	239.479208	0.006281	0.030
rs12568231	239.485489	0.004014	0.032
rs12138275	239.489503	0.003164	0.136
rs2153844	239.492667	0.007038	0.435
rs2580237	239.499705	0.005897	0.492
rs2580223	239.505602	0.003843	0.412
rs2809985	239.509445	0.002559	0.405
rs2810006	239.512004	0.000195	0.263
rs2810007	239.512199	0.000877	0.377

rs924775	239.513076	0.015651	0.183
rs1503793	239.528727	0.001035	0.295
rs7550063	239.529762	0.004911	0.295
rs6657483	239.534673	0.003428	0.304
rs2036493	239.538101	0.000769	0.325
rs1553442	239.538870	0.000483	0.042
rs1039529	239.539353	0.000281	0.413
rs2654875	239.539634	0.000948	0.378
rs1394060	239.540582	0.016479	0.326
rs2810035	239.557061	0.001269	0.398
rs2810038	239.558330	0.007264	0.404
rs2654867	239.565594	0.000070	0.225
rs1473576	239.565664	0.000089	0.371
rs1473575	239.565753	0.000314	0.367
rs1473574	239.566067	0.009125	0.370
rs10926747	239.575192	0.004791	0.443
rs6429357	239.579983	0.002459	0.498
rs905586	239.582442	0.004866	0.450
rs3964039	239.587308	0.000542	0.255
rs1503792	239.587850	0.000797	0.458
rs1548160	239.588647	0.011589	0.459
rs10803047	239.600236	0.001789	0.290
rs6695046	239.602025	0.004642	0.291
rs2809987	239.606667	0.008097	0.446
rs1553435	239.614764	0.005530	0.220
rs2654892	239.620294	0.004008	0.219
rs2654897	239.624302	0.004954	0.325
rs2810005	239.629256	0.000153	0.403
rs2654902	239.629409	0.002713	0.326
rs1503803	239.632122	0.001992	0.334
rs2810008	239.634114	0.019053	0.295
rs12403816	239.653167	0.004091	0.467
rs12121644	239.657258	0.000282	0.033
rs7551410	239.657540	0.000139	0.435
rs1005007	239.657679	0.019716	0.444
rs7544211	239.677395	0.014186	0.116
rs1938338	239.691581	0.041639	0.235
rs6659974	239.733220	0.013504	0.287
rs4658491	239.746724	0.000617	0.011
rs1938336	239.747341	0.006333	0.262
rs6681770	239.753674	0.000780	0.205
rs1333701	239.754454	0.002433	0.136
rs12084486	239.756887	0.001431	0.421
rs2184185	239.758318	0.006778	0.421
rs9804068	239.765096	0.001810	0.284

rs12091230	239.766906	0.005891	0.002
rs1475727	239.772797	0.012538	0.119
rs4658847	239.785335	0.018243	0.070
rs10926825	239.803578	0.010578	0.379
rs1981064	239.814156	0.022060	0.462
rs914940	239.836216	0.054357	0.427
rs7542171	239.890573	0.002792	0.001
rs2095262	239.893365	0.013135	0.299
rs2780784	239.906500	0.038168	0.382
rs2027040	239.944668	0.002902	0.430
rs10926880	239.947570	0.003868	0.001
rs4658862	239.951438	0.016451	0.447
rs7549075	239.967889	0.013428	0.448
rs4639742	239.981317	0.000157	0.007
rs4233436	239.981474	0.014380	0.381
rs10926905	239.995854	0.018773	0.329
rs2502333	240.014627	0.053109	0.224
rs1081093	240.067736	0.013682	0.419
rs1985630	240.081418	0.136759	0.260
rs4658536	240.218177	0.041188	0.140
rs3766666	240.259365	0.016408	0.008
rs3766664	240.275773	0.033483	0.437
rs2789173	240.309256	0.017981	0.337
rs7518350	240.327237	0.010618	0.402
rs4658547	240.337855	0.018158	0.124
rs7548435	240.356013	0.003647	0.016
rs4658552	240.359660	0.003589	0.401
rs3904683	240.363249	0.034430	0.399
rs6682448	240.397679	0.011198	0.360
rs10926984	240.408877	0.009309	0.144
rs976529	240.418186	0.022445	0.027
rs2275155	240.440631	0.019140	0.333
rs10754803	240.459771	0.022213	0.480
rs6695785	240.481984	0.003751	0.430
rs11808637	240.485735	0.018648	0.014
rs10927006	240.504383	0.012871	0.144
rs12120901	240.517254	0.008582	0.017
rs10927011	240.525836	0.027467	0.493
rs7551067	240.553303	0.010609	0.151
rs11810833	240.563912	0.019482	0.175
rs9428573	240.583394	0.015136	0.002
rs12042298	240.598530	0.001633	0.244
rs884808	240.600163	0.003960	0.250
rs3006927	240.604123	0.004628	0.126
rs9428576	240.608751	0.005873	0.479

rs9428966	240.614624	0.019013	0.302
rs4515770	240.633637	0.030679	0.250
rs2220276	240.664316	0.014605	0.324
rs2291409	240.678921	0.029538	0.324
rs320302	240.708459	0.026274	0.127
rs12032481	240.734733	0.018429	0.024
rs2290753	240.753162	0.020647	0.325
rs11578053	240.773809	0.034655	0.031
rs320303	240.808464	0.006658	0.103
rs1486475	240.815122	0.009250	0.280
rs2125229	240.824372	0.000855	0.126
rs2034915	240.825227	0.025863	0.321
rs12124113	240.851090	0.012112	0.153
rs1545654	240.863202	0.043065	0.124
rs6672195	240.906267	0.031290	0.324
rs4518884	240.937557	0.023730	0.322
rs12094774	240.961287	0.009104	0.047
rs3008657	240.970391	0.024195	0.324
rs12069752	240.994586	0.020848	0.001
rs2206	241.015434	0.022606	0.408
rs1568103	241.038040	0.003367	0.335
rs525234	241.041407	0.001364	0.326
rs559080	241.042771	0.000209	0.298
rs471567	241.042980	0.004146	0.422
rs524489	241.047126	0.001387	0.435
rs579388	241.048513	0.004324	0.375
rs990794	241.052837	0.006493	0.260
rs472276	241.059330	0.002312	0.397
rs519449	241.061642	0.000873	0.287
rs10927093	241.062515	0.003420	0.196
rs989993	241.065935	0.001819	0.041
rs1389620	241.067754	0.000311	0.001
rs500202	241.068065	0.002050	0.456
rs484459	241.070115	0.000078	0.295
rs2647328	241.070193	0.010919	0.018
rs511659	241.081112	0.001828	0.281
rs2029742	241.082940	0.001639	0.194
rs495843	241.084579	0.001135	0.486
rs10927094	241.085714	0.000488	0.331
rs960091	241.086202	0.002354	0.183
rs2131817	241.088556	0.003847	0.184
rs546296	241.092403	0.000395	0.158
rs539760	241.092798	0.021057	0.241
rs1009375	241.113855	0.001105	0.254
rs1319790	241.114960		0.182

APPENDIX E

Study Subjects

DNA Samples

To date, the Multiple Sclerosis Genetics Group (MSGG) has collected a data set of 192 multiplex families. The MSGG currently has over 200 families ascertained and awaiting follow-up. The MSGG has also collected a large data set of 593 simplex families with either: a.) at least one affected and both parents or b.) at least one affected and one unaffected sibling that can be used for family-based association studies. The MSGG also has access to independent data sets containing 489 African-American simplex families and 1,000 U.K. parent-child trios that could serve as independent data sets for confirmation of findings (Table 1).

Table 1. Available Data Sets

Family Type	# Families	# Affecteds	# Unaffecteds
Multiplex	192	492	881
Simplex (U.S.)	593	593	1305
Simplex (U.K)	1000	1000	2000
Simplex (Af-Am)	489	489	339

Detailed clinical data was collected for each individual enrolled in the MSGG studies. Each affected individual underwent a detailed medical history, MRI, and lab tests to rule out other forms of neurological disability. A majority of families in the data set exhibit the relapsing-remitting form of MS (93% of multiplex families and 70% of simplex

families). With the exception of the 489 African-American simplex families, all of the available data sets are Caucasian.

Human Subjects

The discussion below describes the approach that the MSGG uses in collection of MS families.

Risks to Human Subjects

Involvement and Characteristics of Human Subjects: Subjects participating in the study either have MS or are related to a person or persons known to have MS. Taking part in MSGG studies is entirely voluntary. Every effort is made to encourage participation of minorities in MSGG research activities, but they constitute a small portion of the overall data set, given the lower frequency of MS in minority populations. The MSGG is specifically targeting African-American individuals for collection; these samples are being examined as part of separate proposals and will potentially be used as a confirmation data set for future association studies. The gender and ethnic distributions of the data set are provided in the Inclusion Enrollment Report.

Sources of Research Material: This information was collected for research purposes only, and will only be accessed in a coded and confidential manner. The vast majority of the data set has been collected in the United States and the information and samples have been collected under separately funded proposals by MSGG members at UCSF. We also

have access to 1,000 parent-child trios collected by MSGG collaborators in the United Kingdom.

Potential Risks: The physical risks of participating in the MSGG studies are minimal, and arise primarily from phlebotomy. These risks included bruising around the site of phlebotomy, and the possibility of passing out from the site of blood. All blood collection was performed by trained personnel who took all necessary precautions to ameliorate such problems. There is a small psychological risk from an individual identifying themselves as part of a family with MS. This is unlikely since most family members participating in these studies are already aware of MS in their family. There is also a small theoretical risk for social and legal discrimination towards individuals who are at risk for a medical disorder or have a medical disorder/condition such as MS in their family.

Protection of Human Subjects

Recruitment and Informed Consent: Informed consent was obtained from all participants prior to enrollment in the MSGG studies. Each participant read the Informed Consent Form, and was given an opportunity to address any questions with the study site senior investigators. If the participant was visually impaired and unable to read the form, study personnel read the form aloud. Participants were given the opportunity to ask questions before signing the informed consent. For cognitively impaired individuals, consent was obtained from a family member and/or legal guardian according to local, state, or federal law. Every effort was made to explain the study to the participant to the best of their

understanding. A signed consent form is stored with research records, and kept completely separate from any clinical medical records. Any participant may refuse to participate and may withdraw consent and discontinue participation in the study at any time without affecting their present or future care.

Database core facilities maintain and update pedigree data information. Study participant confidentiality is maintained through the use of numerically coded samples. In addition to its role in managing the pedigree data, the core facility also databases all genetic marker data. This is done via the PEDIGENE® system to specifically suit genetic research study needs. Multiple security measures are in place to protect data—including data encryption, firewalls, restricted access to the network, and multiple levels of password protection.

There is a theoretical risk for discrimination towards individuals who are at risk for a medical disorder or have a medical disorder/condition in their family. Potential discrimination may include barriers to insurability, employability, or other unidentified adverse effects. Extensive efforts are made to protect all research subjects from prejudice, discrimination, or uses of this information that will adversely affect them. Specifically, clinical and research information with respect to this study is maintained in a research file separate from hospital medical records and will not be placed in the official medical record. Access to information generated as a result of the research study is prohibited and consequently is unavailable to the patient, health care providers,

insurance carriers, employers, or any member of the participant's family. A certificate of confidentiality has been obtained by the MSGG for MS studies.

The collection of patient samples is for research purposes only. Therefore, no results from our studies will be given to the participants. It is possible that these studies will identify information about the participant that was previously unknown (e.g. disease status, risk, or non-paternity). Such incidental findings are not shared with anyone related to the participant unless the incidental finding is life-threatening.

Potential Benefits

Although there is likely to be no direct benefit of this research to the participants, the information generated as part of these studies may help us to better understand the etiology and expression of MS. Participants may benefit from knowing that their participation may help us to better understand MS and may eventually lead to better treatments for themselves and/or others. The risks of participation are minimal and the potential benefits to society are substantial.

Importance of the Knowledge to be Gained

Identifying one or more genes for MS risk or disease expression may provide improved diagnosis and better discrimination of the subtypes of MS. The genes may also provide substrates for more effective treatments (e.g. better pharmacological agents) or better targeting of current treatments to those most likely to respond.

Inclusion of Women and Minorities

MS occurs in women more often than in men (~ 2.5:1) and is more common in Caucasian populations (particularly of northern European descent) than in other minority populations. Our study population consists of more females than males and includes a substantial number of non-Caucasian samples. Approximate gender and ethnic distributions of samples to be used in this study are shown in the Inclusion Enrollment Report. For this study, we will concentrate our efforts on the Caucasian data set, which represents the vast majority of all collected samples. Any findings identified through this proposal will specifically be tested in the African-American data set, and any other minority data set (e.g. Hispanics) of sufficient size to be scientifically valid.

Inclusion of Children

Multiple sclerosis is a disease of young to mid-adult life. Few individuals under the age of 18 are affected with the disease and therefore few such individuals are included in our studies. The incidence of MS increases after the age of 18, so individuals between the ages of 18 and 21 (still considered children by NIH definition) are to be included if they are affected. Unaffected children under the age of 18 will not be included. Affected children over the age of 10 will be included. It is unlikely that sufficient numbers of affected children will be included to make specific statements about any differences in their genetic risk.

REFERENCES

1. Genain, C.P., Cannella, B., Hauser, S.L., Raine, C.S. (1999) Identification of autoantibodies associated with myelin damage in multiple sclerosis. *Nat Med*, **5**, 170-175.

2. Raine, C.S., Cannella, B., Hauser, S.L., Genain, C.P. (1999) Demyelination in primate autoimmune encephalomyelitis and acute multiple sclerosis lesions: a case for antigen-specific antibody mediation. *Ann Neurol*, **46**, 144-160.

3. Lucchinetti, C., Bruck, W., Parisi, J., Scheithauer, B., Rodriquez, M., Lassmann, H. (2000) Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *American Neurological Association*, **47 no 6**, 707-717.

4. Hauser, S.L., Goodkin, D.E. (1998) In Fauci, A., Braunwald, E., Isselbacher, J., Martin, J., Kasper, D., Hauser, S.L., Longo, D. (eds.), *Harrison's Principle of Internal Medicine*. McGraw Hill, New York, pp. 2409-2419.

5. Goodkin, D.E., Doolittle, T.H., Hauser, S.S., Ransohoff, R.M., Roses, A.D. (1991) Diagnostic criteria for multiple sclerosis research involving multiply affected families. *Arch Neurology*, **48**, 805-807.

6. Barcellos, L.F., Oksenberg, J.R., Green, A.J., Bucher, P., Rimmler, J.B., Schmidt, S., Garcia, M.E., Lincoln, R.R., Pericak-Vance, M.A., Haines, J.L., Hauser, S.L. (2002) Genetic basis for clinical expression in multiple sclerosis. *Brain*, **125**, 150-158.

7. Cook, S.D., MacDonald, J., Tapp, W., Poskanzer, D., Dowling, P.C. (1988) Multiple sclerosis in the Shetland Islands: an update. *Acta Neurol Scand*, **77**, 148-151.

8. Yu, Y.L., Woo, E., Hawkins, B.R., Ho, H.C., Huang, C.Y. (1989) Multiple sclerosis amongst Chinese in Hong Kong. *Brain*, **112 (Pt 6)**, 1445-1467.

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

10. Sadovnick, A.D., Armstrong, H., Rice, G., Bulman, D.E., Hashimoto, L., Paty, D.W., Hashimoto, S., Warren, S., Hader, W., Murray, T.J. (1993) A population based study of multiple sclerosis in twins: update. *Ann Neurol*, **33**, 281-285.

11. Mumford, G.L., Wood, N.W., Kellar-Wood, H.F., Thorpe, J.W., Miler, D.H. (1994) The British Isles survey of multiple sclerosis in twins. *Neurology*, **44**, 15.

12. Sadovnick, A.D., Ebers, G.C., Dyment, D.A., Risch, N.J. (1996) Evidence for genetic basis of multiple sclerosis. *Lancet*, **347**, 1730.

13. Robertson, N.P., Fraser, M., Deans, J., Clayton, D., Walker, N., Compston, D.A. (1996) Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. *Brain*, **119 (Pt 2)**, 449-455.

14. Sadovnick, A.D., Ebers, G.C. (1995) Genetics of multiple sclerosis. *Neurologic Clinics*, **13**, 99-118.

15. Ebers, G.C., Sadovnick, A.D., Risch, N.J. (1995) A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group [see comments]. *Nature*, **377**, 150-151.

16. Pericak-Vance, M.A., Haines, J.L. (1995) Genetic susceptibility to Alzheimer disease. *Trends in Genetics*, **11**, 504-508.

17. Polymeropoulos, M.H., Higgins, J.J., Golbe, L.I., Johnson, W.G., Ide, S.E., Di Iorio, G., Sanges, G., Stenroose, E.S., Pho, L.T., Schaffer, A.A.*, et al.* (1996) Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science*, **274**, 1197-1199.

18. Scott, W.K., Stajich, J.M., Yamaoka, L.H., Speer, M.C., Vance, J.M., Roses, A.D., Pericak-Vance, M.A., Deane Laboratory Parkinson Research Group (1997) Genetic complexity and Parkinson's disease. *Science*, **277**, 387-389.

19. Pericak-Vance, M.A., Grubber, J., Bailey, L.R., Hedges, D., West, S., Santoro, L., Kemmerer, B., Hall, J.L., Saunders, A.M., Roses, A.D.*, et al.* (2000) Identification of novel genes in late-onset Alzheimer's disease. *Exp Gerontol*, **35**, 1343-1352.

20. Scott, W.K., Nance, M.A., Watts, R.L., Hubble, J.P., Koller, W.C., Lyons, K., Pahwa, R., Stern, M.B., Colcher, A., Hiner, B.C.*, et al.* (2001) Complete genomic screen in Parkinson disease: evidence for multiple genes. *JAMA*, **286**, 2239-2244.

21. Bertrams, J., Kuwert, E. (1972) HL-A antigen frequencies in multiple sclerosis. Significant increase of HL-A3, HL-A10 and W5, and decrease of HL-A12. *Eur J Neurol*, **7**, 78.

22. Naito, S., Namerow, N., Mickey, M.R., Terasaki, P.I. (1972) Multiple sclerosis: association with HL-A3. *Tissue Antigens*, **2**, 1-4.

23. Compston, A., Batchelor, J.R., McDonald, W.I. (1976) B-lymphocyte alloantigens associated with multiple sclerosis. *Lancet*, **2**, 1261-1265.

24. Haines, J.L., Terwedow, H.A., Burgess, K., Pericak-Vance, M.A., Rimmler, J.B., Martin, E.R., Oksenberg, J.R., Lincoln, R., Zhang, D.Y., Banatao, D.R.*, et al.* (1998) Linkage of the MHC to familial multiple sclerosis suggests genetic

heterogeneity. The Multiple Sclerosis Genetics Group. *Hum Mol Genet*, **7**, 1229-1234.

25. Oksenberg, J.R., Barcellos, L.F., Cree, B.A., Baranzini, S.E., Bugawan, T.L., Khan, O., Lincoln, R.R., Swerdlin, A., Mignot, E., Lin, L.*, et al.* (2004) Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. *Am J Hum Genet*, **74**, 160-167.

26. Tienari, P.J., Wikstrom, J., Sajantila, A., Palo, J., Peltonen, L. (1992) Genetic susceptibility to multiple sclerosis is linked to myelin basic protein gene. *Lancet*, **340**, 987-991.

27. Graham, C.A., Kirk, C.W., Nevin, N.C., Droogan, A.G., Hawkins, S.A., McMillan, S.A., McNeill, T.A. (1993) Lack of association between myelin basic protein gene microsatellite and multiple sclerosis. *Lancet*, **341**, 1596.

28. Rose, J., Gerken, S., Lynch, S., Pisani, P., Varvil, T., Otterud, B., Leppert, M. (1993) Genetic susceptibility in familial multiple sclerosis not linked to the myelin basic protein gene. *Lancet*, **341**, 1179-1181.

29. Eoli, M., Pandolfo, M., Milanese, C., Gasparini, P., Salmaggi, A., Zeviani, M. (1994) The myelin basic protein gene is not a major susceptibility locus for multiple sclerosis in Italian patients. *J Neurol*, **241**, 615-619.

30. Vandevyver, C., Stinissen, P., Cassiman, J.J., Raus, J. (1994) Myelin basic protein gene polymorphism is not associated with chronic progressive multiple sclerosis. *J Neuroimmunol*, **52**, 97-99.

31. Wood, N.W., Holmans, P., Clayton, D., Robertson, N., Compston, D.A. (1994) No linkage or association between multiple sclerosis and the myelin basic protein gene in affected sibling pairs. *J Neurol Neurosurg Psychiatry*, **57**, 1191-1194.

32. Price, S.E., Sharpe, G., Boots, A., Poutsma, A., Mason, C., James, J., Hinks, L., Thompson, R.J. (1997) Role of myelin basic protein and proteolipid protein genes in multiple sclerosis: single strand conformation polymorphism analysis of the human sequences. *Neuropathol Appl Neurobiol*, **23**, 457-467.

33. He, B., Yang, B., Lundahl, J., Fredrikson, S., Hillert, J. (1998) The myelin basic protein gene in multiple sclerosis: identification of discrete alleles of a 1.3 kb tetranucleotide repeat sequence. *Acta Neurol Scand*, **97**, 46-51.

34. Tefferi, A., Bolander, M.E., Ansell, S.M., Wieben, E.D., Spelsberg, T.C. (2002) Primer on medical genomics. Part III: Microarray experiments and data analysis. *Mayo Clin Proc*, **77**, 927-940.

35. Hanash, S. (2003) Disease proteomics. *Nature*, **422**, 226-232.

36. Kronick, M.N. (2004) Creation of the whole human genome microarray. *Expert Rev Proteomics*, **1**, 19-28.

37. Lueking, A., Cahill, D.J., Mullner, S. (2005) Protein biochips: A new and versatile platform technology for molecular medicine. *Drug Discov Today*, **10**, 789-794.

38. Ebers, G.C., Kukay, K., Bulman, D.E., Sadovnick, A.D., Rice, G., Anderson, C., Armstrong, H., Cousin, K., Bell, R.B., Hader, W.*, et al.* (1996) A full genome search in multiple sclerosis. *Nat Genet*, **13**, 472-476.

39. Haines, J.L., Ter Minassian, M., Bazyk, A., Gusella, J.F., Kim, D.J., Terwedow, H., Pericak-Vance, M.A., Rimmler, J.B., Haynes, C.S., Roses, A.D.*, et al.* (1996) A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. The Multiple Sclerosis Genetics Group. *Nat Genet*, **13**, 469-471.

40. Sawcer, S., Jones, H.B., Feakes, R., Gray, J., Smaldon, N., Chataway, J., Robertson, N., Clayton, D., Goodfellow, P.N., Compston, A. (1996) A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat Genet*, **13**, 464-468.

41. Kuokkanen, S., Gschwend, M., Rioux, J.D., Daly, M.J., Terwilliger, J.D., Tienari, P.J., Wikstrom, J., Palo, J., Stein, L.D., Hudson, T.J.*, et al.* (1997) Genomewide scan of multiple sclerosis Finnish multiplex families. *Am J Hum Genet*, **61**, 1379-1387.

42. Chataway, J., Feakes, R., Coraddu, F., Gray, J., Deans, J., Fraser, M., Robertson, N., Broadley, S., Jones, H., Clayton, D.*, et al.* (1998) The genetics of multiple sclerosis: principles, background and updated results of the United Kingdom systematic genome screen. *Brain*, **121 (Pt 10)**, 1869-1887.

43. Wise, L.H., Lanchbury, J.S., Lewis, C.M. (1999) Meta-analysis of genome searches. *Ann Hum Genet*, **63 (Pt 3)**, 263-272.

44. Broadley, S., Sawcer, S., D'Alfonso, S., Hensiek, A., Coraddu, F., Gray, J., Roxburgh, R., Clayton, D., Buttinelli, C., Quattrone, A.*, et al.* (2001) A genome screen for multiple sclerosis in Italian families. *Genes Immun*, **2**, 205-210.

45. Coraddu, F., Sawcer, S., D'Alfonso, S., Lai, M., Hensiek, A., Solla, E., Broadley, S., Mancosu, C., Pugliatti, M., Marrosu, M.G., Compston, A. (2001) A genome screen for multiple sclerosis in Sardinian multiplex families. *Eur J Hum Genet*, **9**, 621-626.

46. Akesson, E., Oturai, A., Berg, J., Fredrikson, S., Andersen, O., Harbo, H.F., Laaksonen, M., Myhr, K.M., Nyland, H.I., Ryder, L.P.*, et al.* (2002) A genome-wide screen for linkage in Nordic sib-pairs with multiple sclerosis. *Genes Immun*, **3**, 279-285.

47. Ban, M., Stewart, G.J., Bennetts, B.H., Heard, R., Simmons, R., Maranian, M., Compston, A., Sawcer, S.J. (2002) A genome screen for linkage in Australian sibling-pairs with multiple sclerosis. *Genes Immun*, **3**, 464-469.

48. Haines, J.L., Bradford, Y., Garcia, M.E., Reed, A.D., Neumeister, E., Pericak-Vance, M.A., Rimmler, J.B., Menold, M.M., Martin, E.R., Oksenberg, J.R.*, et al.* (2002) Multiple susceptibility loci for multiple sclerosis. *Hum Mol Genet*, **11**, 2251-2256.

49. Eraksoy, M., Kurtuncu, M., Akman-Demir, G., Kilinc, M., Gedizlioglu, M., Mirza, M., Anlar, O., Kutlu, C., Demirkiran, M., Idrisoglu, H.A.*, et al.* (2003) A whole genome screen for linkage in Turkish multiple sclerosis. *J Neuroimmunol*, **143**, 17-24.

50. Risch, N. (1987) Assessing the role of HLA-linked and unlinked determinants of disease. *Am J Hum Genet*, **40**, 1-14.

51. Ligers, A., Dyment, D.A., Willer, C.J., Sadovnick, A.D., Ebers, G., Risch, N., Hillert, J. (2001) Evidence of linkage with HLA-DR in DRB1*15-negative families with multiple sclerosis. *Am J Hum Genet*, **69**, 900-903.

52. Sawcer, S., Maranian, M., Setakis, E., Curwen, V., Akesson, E., Hensiek, A., Coraddu, F., Roxburgh, R., Sawcer, D., Gray, J.*, et al.* (2002) A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility. *Brain*, **125**, 1337-1347.

53. Goedde, R., Sawcer, S., Boehringer, S., Miterski, B., Sindern, E., Haupts, M., Schimrigk, S., Compston, A., Epplen, J.T. (2002) A genome screen for linkage disequilibrium in HLA-DRB1*15-positive Germans with multiple sclerosis based on 4666 microsatellite markers. *Hum Genet*, **111**, 270-277.

54. Ban, M., Sawcer, S.J., Heard, R.N., Bennetts, B.H., Adams, S., Booth, D., Perich, V., Setakis, E., Compston, A., Stewart, G.J. (2003) A genome-wide screen for linkage disequilibrium in Australian HLA-DRB1*1501 positive multiple sclerosis patients. *J Neuroimmunol*, **143**, 60-64.

55. Bielecki, B., Mycko, M.P., Tronczynska, E., Bieniek, M., Sawcer, S., Setakis, E., Benediktsson, K., Compston, A., Selmaj, K.W. (2003) A whole genome screen for association in Polish multiple sclerosis patients. *J Neuroimmunol*, **143**, 107-111.

56. Coraddu, F., Lai, M., Mancosu, C., Cocco, E., Sawcer, S., Setakis, E., Compston, A., Marrosu, M.G. (2003) A genome-wide screen for linkage disequilibrium in Sardinian multiple sclerosis. *J Neuroimmunol*, **143**, 120-123.

57. Eraksoy, M., Hensiek, A., Kurtuncu, M., Kman-Demir, G., Kilinc, M., Gedizlioglu, M., Petek-Balci, B., Anlar, O., Kutlu, C., Saruhan-Direskeneli, G.*, et*

al. (2003) A genome screen for linkage disequilibrium in Turkish multiple sclerosis. *J Neuroimmunol*, **143**, 129-132.

58. Goertsches, R., Villoslada, P., Comabella, M., Montalban, X., Navarro, A., de la Concha, E.G., Arroyo, R., Lopez de, M.A., Otaegui, D., Palacios, R.*, et al.* (2003) A genomic screen of Spanish multiple sclerosis patients reveals multiple loci associated with the disease. *J Neuroimmunol*, **143**, 124-128.

59. Goris, A., Sawcer, S., Vandenbroeck, K., Carton, H., Billiau, A., Setakis, E., Compston, A., Dubois, B. (2003) New candidate loci for multiple sclerosis susceptibility revealed by a whole genome association screen in a Belgian population. *J Neuroimmunol*, **143**, 65-69.

60. Harbo, H.F., Datta, P., Oturai, A., Ryder, L.P., Sawcer, S., Setakis, E., Akesson, E., Celius, E.G., Modin, H., Sandberg-Wollheim, M.*, et al.* (2003) Two genome-wide linkage disequilibrium screens in Scandinavian multiple sclerosis patients. *J Neuroimmunol*, **143**, 101-106.

61. Heggarty, S., Sawcer, S., Hawkins, S., McDonnell, G., Droogan, A., Vandenbroeck, K., Hutchinson, M., Setakis, E., Compston, A., Graham, C. (2003) A genome wide scan for association with multiple sclerosis in a N. Irish case control population. *J Neuroimmunol*, **143**, 93-96.

62. Liguori, M., Sawcer, S., Setakis, E., Compston, A., Giordano, M., D'Alfonso, S., Mellai, M., Malferrari, G., Trojano, M., Livrea, P.*, et al.* (2003) A whole genome screen for linkage disequilibrium in multiple sclerosis performed in a continental Italian population. *J Neuroimmunol*, **143**, 97-100.

63. Rajda, C., Bencsik, K., Seres, E., Jonasdottir, A., Foltynie, T., Sawcer, S., Benediktsson, K., Fossdal, R., Setakis, E., Compston, A., Vecsei, L. (2003) A genome-wide screen for association in Hungarian multiple sclerosis. *J Neuroimmunol*, **143**, 84-87.

64. Santos, M., Pinto-Basto, J., Rio, M.E., Sa, M.J., Valenca, A., Sa, A., Dinis, J., Figueiredo, J., Bigotte de, A.L., Coelho, I.*, et al.* (2003) A whole genome screen for association with multiple sclerosis in Portuguese patients. *J Neuroimmunol*, **143**, 112-115.

65. Weber, A., Infante-Duarte, C., Sawcer, S., Setakis, E., Bellmann-Strobl, J., Hensiek, A., Rueckert, S., Schoenemann, C., Benediktsson, K., Compston, A., Zipp, F. (2003) A genome-wide German screen for linkage disequilibrium in multiple sclerosis. *J Neuroimmunol*, **143**, 79-83.

66. Sawcer, S., Ban, M., Maranian, M., Yeo, T.W., Compston, A., Kirby, A., Daly, M.J., De Jager, P.L., Walsh, E., Lander, E.S.*, et al.* (2005) A high-density screen for linkage in multiple sclerosis. *Am J Hum Genet*, **77**, 454-467.

67. Chapman, J., Sylantiev, C., Nisipeanu, P., Korczyn, A.D. (1999) Preliminary observations on APOE epsilon4 allele and progression of disability in multiple sclerosis. *Arch Neurol*, **56**, 1484-1487.

68. Evangelou, N., Jackson, M., Beeson, D., Palace, J. (1999) Association of the APOE epsilon4 allele with disease activity in multiple sclerosis. *J Neurol Neurosurg Psychiatry*, **67**, 203-205.

69. D'Alfonso, S., Nistico, L., Bocchio, D., Bomprezzi, R., Marrosu, M.G., Murru, M.R., Lai, M., Massacesi, L., Ballerini, C., Repice, A.*, et al.* (2000) An attempt of identifying MS-associated loci as a follow-up of a genomic linkage study in the Italian population. *J Neurovirol*, **6 Suppl 2**, S18-S22.

70. Fazekas, F., Strasser-Fuchs, S., Kollegger, H., Berger, T., Kristoferitsch, W., Schmidt, H., Enzinger, C., Schiefermeier, M., Schwarz, C., Kornek, B.*, et al.* (2001) Apolipoprotein E epsilon 4 is associated with rapid progression of multiple sclerosis. *Neurology*, **57**, 853-857.

71. Fazekas, F., Strasser-Fuchs, S., Schmidt, H., Enzinger, C., Ropele, S., Lechner, A., Flooh, E., Schmidt, R., Hartung, H.P. (2000) Apolipoprotein E genotype related differences in brain lesions of multiple sclerosis. *J Neurol Neurosurg Psychiatry*, **69**, 25-28.

72. Chapman, J., Vinokurov, S., Achiron, A., Karussis, D.M., Mitosek-Szewczyk, K., Birnbaum, M., Michaelson, D.M., Korczyn, A.D. (2001) APOE genotype is a major predictor of long-term progression of disability in MS. *Neurology*, **56**, 312-316.

73. Hogh, P., Oturai, A., Schreiber, K., Blinkenberg, M., Jorgensen, O.S., Ryder, L., Paulson, O.B., Sorensen, P.S., Knudsen, G.M. (2000) Apoliprotein E and multiple sclerosis: impact of the epsilon-4 allele on susceptibility, clinical type and progression rate. *Mult Scler*, **6**, 226-230.

74. Schmidt, S., Barcellos, L.F., DeSombre, K., Rimmler, J.B., Lincoln, R.R., Bucher, P., Saunders, A.M., Lai, E., Martin, E.R., Vance, J.M.*, et al.* (2002) Association of polymorphisms in the apolipoprotein e region with susceptibility to and progression of multiple sclerosis. *Am J Hum Genet*, **70**, 708-717.

75. Tienari, P.J., Reunanen, K. (1999) Chromosome 19 locus apolipoprotein C-II association with multiple sclerosis. *Mult Scler*, **5**, 378.

76. Kirk, C.W., Graham, C.A., McDonnell, G.V., Hawkins, S.A. (2000) Chromosome 19 locus apolipoprotein C-II association with multiple sclerosis. *Mult Scler*, **6**, 291-292.

77. Zouali, H., Faure-Delanef, L., Lucotte, G. (1999) Chromosome 19 locus apolipoprotein C-II association with multiple sclerosis. *Mult Scler*, **5**, 134-136.

78. Koch, S., Goedde, R., Nigmatova, V., Epplen, J.T., Muller, N., de, S.J., Vermersch, P., Momot, T., Schmidt, R.E., Witte, T. (2005) Association of multiple sclerosis with ILT6 deficiency. *Genes Immun*, **6**, 445-447.

79. Spielman, R.S., McGinnis, R.E., Ewens, W.J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*, **52**, 506-516.

80. Spielman, R.S., Ewens, W.J. (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am J Hum Genet*, **61**, 450-458.

81. Martin, E.R., Monks, S.A., Warren, L.L., Kaplan, N.L. (2000) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet*, **67**, 146-154.

82. Martin, E.R., Bass, M.P., Gilbert, J.R., Pericak-Vance, M.A., Hauser, E.R. (2003) Genotype-based association test for general pedigrees: the genotype-PDT. *Genet Epidemiol*, **25**, 203-213.

83. Hutcheson, H.B., Bradford, Y., Folstein, S.E., Gardiner, M.B., Santangelo, S., Sutcliffe, J.S., Haines, J.L. (2003) Defining the Autism Minimum Candidate Gene Region on Chromosome 7. *Am J Med Genet*, **117B**, 90-96.

84. Cottingham, R.W. Jr., Idury, R.M., Schaffer, A.A. (1993) Faster sequential genetic linkage computations. *Am J Hum Genet*, **53**, 252-263.

85. Schaffer, A.A., Gupta, S.K., Shriram, K., Cottingham, R.W. (1994) Avoiding recomputation in linkage analysis. *Hum Hered*, **44**, 225-237.

86. Smith, C.A.B. (1963) Testing for heterogeneity of recombination fractions values in human genetics. *Ann Hum Genet*, **27**, 175-182.

87. Ott, J. (1986) The number of families required to detect or exclude linkage heterogeneity. *Am J Hum Genet*, **39**, 159-165.

88. Sobel, E., Lange, K. (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet*, **58**, 1323-1337.

89. Vance, J.M. (1997) In Haines, J.L., Pericak-Vance, M.A. (eds.), *Design and Implementation of Mapping Studies for Complex Human Diseases.* Wiley and Sons, New York.

90. O'Connell, J.R., Weeks, D.E. (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet*, **63**, 259-266.

91. Epstein, M.P., Duren, W.L., Boehnke, M. (2000) Improved inference of relationship for pairs of individuals. *Am J Hum Genet*, **67**, 1219-1231.

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

93. Gudbjartsson, D.F., Jonasson, K., Frigge, M.L., Kong, A. (2000) Allegro, a new computer program for multipoint linkage analysis. *Nat Genet*, **25**, 12-13.

94. Laaksonen, M., Jonasdottir, A., Fossdal, R., Ruutiainen, J., Sawcer, S., Compston, A., Benediktsson, K., Thorlacius, T., Gulcher, J., Ilonen, J. (2003) A whole genome association study in Finnish multiple sclerosis patients with 3669 markers. *J Neuroimmunol*, **143**, 70-73.

95. Jawaheer, D., Seldin, M.F., Amos, C.I., Chen, W.V., Shigeta, R., Monteiro, J., Kern, M., Criswell, L.A., Albani, S., Nelson, J.L.*, et al.* (2001) A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. *Am J Hum Genet*, **68**, 927-936.

96. Shai, R., Quismorio, F.P., Jr., Li, L., Kwon, O.J., Morrison, J., Wallace, D.J., Neuwelt, C.M., Brautbar, C., Gauderman, W.J., Jacob, C.O. (1999) Genome-wide screen for systemic lupus erythematosus susceptibility genes in multiplex families. *Hum Mol Genet*, **8**, 639-644.

97. Roth, M.P., Viratelle, C., Dolbois, L., Delverdier, M., Borot, N., Pelletier, L., Druet, P., Clanet, M., Coppin, H. (1999) A genome-wide search identifies two susceptibility loci for experimental autoimmune encephalomyelitis on rat chromosomes 4 and 10. *J Immunol*, **162**, 1917-1922.

98. Rioux, J.D., Daly, M.J., Silverberg, M.S., Lindblad, K., Steinhart, H., Cohen, Z., Delmonte, T., Kocher, K., Miller, K., Guschwan, S.*, et al.* (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet*, **29**, 223-228.

99. Morahan, G., Huang, D., Ymer, S.I., Cancilla, M.R., Stephen, K., Dabadghao, P., Werther, G., Tait, B.D., Harrison, L.C., Colman, P.G. (2001) Linkage disequilibrium of a type 1 diabetes susceptibility locus with a regulatory IL12B allele. *Nat Genet*, **27**, 218-221.

100. Greco, L., Corazza, G., Babron, M.C., Clot, F., Fulchignoni-Lataud, M.C., Percopo, S., Zavattari, P., Bouguerra, F., Dib, C., Tosi, R.*, et al.* (1998) Genome search in celiac disease. *Am J Hum Genet*, **62**, 669-675.

101. Naluai, A.T., Nilsson, S., Gudjonsdottir, A.H., Louka, A.S., Ascher, H., Ek, J., Hallberg, B., Samuelsson, L., Kristiansson, B., Martinsson, T.*, et al.* (2001) Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for celiac disease on chromosomes 5 and 11. *Eur J Hum Genet*, **9**, 938-944.

102. Cookson, W.O., Moffatt, M.F. (2000) Genetics of asthma and allergic disease. *Hum Mol Genet*, **9**, 2359-2364.

103. Enzinger, C., Ropele, S., Smith, S., Strasser-Fuchs, S., Poltrum, B., Schmidt, H., Matthews, P.M., Fazekas, F. (2004) Accelerated evolution of brain atrophy and "black holes" in MS patients with APOE-epsilon 4. *Ann Neurol*, **55**, 563-569.

104. Leal, S.M., Ott, J. (2000) Effects of stratification in the analysis of affected-sib-pair data: benefits and costs. *Am J Hum Genet*, **66**, 567-575.

105. Kenealy, S.J., Babron, M.C., Bradford, Y., Schnetz-Boutaud, N., Haines, J.L., Rimmler, J.B., Schmidt, S., Pericak-Vance, M.A., Barcellos, L.F., Lincoln, R.R.*, et al.* (2004) A second-generation genomic screen for multiple sclerosis. *Am J Hum Genet*, **75**, 1070-1078.

106. Poser, C.M., Paty, D.W., Scheinberg, L., McDonald, W.I., Davis, F.A., Ebers, G.C., Johnson, K.P., Sibley, W.A., Silberberg, D.H., Tourtellotte, W.W. (1983) New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol*, **13**, 227-231.

107. Hauser, E.R., Watanabe, R.M., Duren, W.L., Bass, M.P., Langefeld, C.D., Boehnke, M. (2004) Ordered subset analysis in genetic linkage mapping of complex traits. *Genet Epidemiol*, **27**, 53-63.

108. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P., Lander, E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet*, **58**, 1347-1363.

109. Sawcer, S.J., Maranian, M., Singlehurst, S., Yeo, T., Compston, A., Daly, M.J., De Jager, P.L., Gabriel, S., Hafler, D.A., Ivinson, A.J.*, et al.* (2004) Enhancing linkage analysis of complex disorders: an evaluation of high-density genotyping. *Hum Mol Genet*, **13**, 1943-1949.

110. MacKay, K., Eyre, S., Myerscough, A., Milicic, A., Barton, A., Laval, S., Barrett, J., Lee, D., White, S., John, S.*, et al.* (2002) Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. *Arthritis Rheum*, **46**, 632-639.

111. Jawaheer, D., Seldin, M.F., Amos, C.I., Chen, W.V., Shigeta, R., Etzel, C., Damle, A., Xiao, X., Chen, D., Lum, R.F.*, et al.* (2003) Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. *Arthritis Rheum*, **48**, 906-916.

112. Gaffney, P.M., Kearns, G.M., Shark, K.B., Ortmann, W.A., Selby, S.A., Malmgren, M.L., Rohlf, K.E., Ockenden, T.C., Messner, R.P., King, R.A.*, et al.* (1998) A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. *Proc Natl Acad Sci U. S. A.*, **95**, 14875-14879.

113. Moser, K.L., Neas, B.R., Salmon, J.E., Yu, H., Gray-McGuire, C., Asundi, N., Bruner, G.R., Fox, J., Kelly, J., Henshall, S.*, et al.* (1998) Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. *Proc Natl Acad Sci U. S. A.*, **95**, 14869-14874.

114. Gaffney, P.M., Ortmann, W.A., Selby, S.A., Shark, K.B., Ockenden, T.C., Rohlf, K.E., Walgrave, N.L., Boyum, W.P., Malmgren, M.L., Miller, M.E.*, et al.* (2000) Genome screening in human systemic lupus erythematosus: results from a second Minnesota cohort and combined analyses of 187 sib-pair families. *Am J Hum Genet*, **66**, 547-556.

115. Scott, W.K., Hauser, E.R., Schmechel, D.E., Welsh-Bohmer, K.A., Small, G.W., Roses, A.D., Saunders, A.M., Gilbert, J.R., Vance, J.M., Haines, J.L., Pericak-Vance, M.A. (2003) Ordered-subsets linkage analysis detects novel Alzheimer disease loci on chromosomes 2q34 and 15q22. *Am J Hum Genet*, **73**, 1041-1051.

116. Shao, Y., Cuccaro, M.L., Hauser, E.R., Raiford, K.L., Menold, M.M., Wolpert, C.M., Ravan, S.A., Elston, L., Decena, K., Donnelly, S.L.*, et al.* (2003) Fine Mapping of Autistic Disorder to Chromosome 15q11-q13 by Use of Phenotypic Subtypes. *Am J Hum Genet*, **72**, 539-548.

117. Schmidt, S., Scott, W.K., Postel, E.A., Agarwal, A., Hauser, E.R., De La Paz, M.A., Gilbert, J.R., Weeks, D.E., Gorin, M.B., Haines, J.L., Pericak-Vance, M.A. (2004) Ordered subset linkage analysis supports a susceptibility locus for age-related macular degeneration on chromosome 16p12. *BMC Genet*, **5**, 18.

118. Weeks, D.E., Conley, Y.P., Tsai, H.J., Mah, T.S., Schmidt, S., Postel, E.A., Agarwal, A., Haines, J.L., Pericak-Vance, M.A., Rosenfeld, P.J.*, et al.* (2004) Age-related maculopathy: a genomewide scan with continued evidence of susceptibility loci within the 1q31, 10q26, and 17q25 regions. *Am J Hum Genet*, **75**, 174-189.

119. Sale, M.M., Freedman, B.I., Langefeld, C.D., Williams, A.H., Hicks, P.J., Colicigno, C.J., Beck, S.R., Brown, W.M., Rich, S.S., Bowden, D.W. (2004) A genome-wide scan for type 2 diabetes in African-American families reveals evidence for a locus on chromosome 6q. *Diabetes*, **53**, 830-837.

120. Bowden, D.W., Colicigno, C.J., Langefeld, C.D., Sale, M.M., Williams, A., Anderson, P.J., Rich, S.S., Freedman, B.I. (2004) A genome scan for diabetic nephropathy in African Americans. *Kidney Int*, **66**, 1517-1526.

121. Xu, J., Langefeld, C.D., Zheng, S.L., Gillanders, E.M., Chang, B.L., Isaacs, S.D., Williams, A.H., Wiley, K.E., Dimitrov, L., Meyers, D.A.*, et al.* (2004) Interaction effect of PTEN and CDKN1B chromosomal regions on prostate cancer linkage. *Hum Genet*, **115**, 255-262.

122. Hauser, E.R., Boehnke, M. (1998) Genetic linkage analysis of complex genetic traits using affected sibling pairs. *Biometrics*, **54**, 1238-1246.

123. Hauser, E.R., Bass, M.P., Martin, E.R., Watanabe, R.M., Duren, W.L., Boehnke, M. (2001) p. A529.

124. Hauser, M.A., Li, Y.J., Takeuchi, S., Walters, R., Noureddine, M., Maready, M., Darden, T., Hulette, C., Martin, E., Hauser, E.*, et al.* (2003) Genomic convergence: identifying candidate genes for Parkinson's disease by combining serial analysis of gene expression and genetic linkage. *Hum Mol Genet*, **12**, 671-677.

125. Li, Y.J., Oliveira, S.A., Xu, P., Martin, E.R., Stenger, J.E., Scherzer, C.R., Hauser, M.A., Scott, W.K., Small, G.W., Nance, M.A.*, et al.* (2003) Glutathione S-Transferase Omega 1 modifies age-at-onset of Alzheimer Disease and Parkinson Disease. *Hum Mol Genet*, **12**, 3259-3267.

126. Loots, G.G., Locksley, R.M., Blankespoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M., Frazer, K.A. (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*, **288**, 136-140.

127. Mortlock, D.P., Guenther, C., Kingsley, D.M. (2003) A general approach for identifying distant regulatory elements applied to the Gdf6 gene. *Genome Res*, **13**, 2069-2081.

128. Nobrega, M.A., Ovcharenko, I., Afzal, V., Rubin, E.M. (2003) Scanning human gene deserts for long-range enhancers. *Science*, **302**, 413.

129. Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., Dubchak, I. (2000) VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics*, **16**, 1046-1047.

130. Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., Miller, W. (2000) PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res*, **10**, 577-586.

131. Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520-562.

132. Dubchak, I., Brudno, M., Loots, G.G., Pachter, L., Mayor, C., Rubin, E.M., Frazer, K.A. (2000) Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res*, **10**, 1304-1306.

133. Margulies, E.H., Blanchette, M., Haussler, D., Green, E.D. (2003) Identification and characterization of multi-species conserved sequences. *Genome Res*, **13**, 2507-2518.

134. Weeks, D.E., Sobel, E., O'Connell, J.R., Lange, K. (1995) Computer programs for multilocus haplotyping of general pedigrees. *Am J Hum Genet*, **56**, 1506-1507.

135. Martin, E.R., Hardy, S.W., Bass, M.P., Kaplan, N.L. (2001) Sampling considerations for family-based tests of association. *Am J Hum Genet*, **67**, 312.

136. Lange, C., DeMeo, D., Silverman, E.K., Weiss, S.T., Laird, N.M. (2004) PBAT: tools for family-based association studies. *Am J Hum Genet*, **74**, 367-369.

137. Barrett, J.C., Fry, B., Maller, J., Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263-265.